

**REGRESSION MODELS FOR DYNAMIC
TREATMENT REGIMENS AND QUANTILE
ASSOCIATION OF BIVARIATE SURVIVAL DATA**

by

Ling-Wan Chen

B.S. in Applied Mathematics, National Dong Hwa University,
Taiwan, 2008

M.S. in Statistics, National Tsing Hua University, Taiwan, 2010

Submitted to the Graduate Faculty of
the Kenneth P. Dietrich School of Arts and Sciences in partial
fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH
THE DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Ling-Wan Chen

It was defended on

July 19, 2018

and approved by

Yu Cheng, Ph.D, Statistics and Biostatistics

Kehui Chen, Ph.D, Statistics

Satish Iyengar, Ph.D, Statistics

Abdus Wahed, Ph.D, Biostatistics

Dissertation Director: Yu Cheng, Ph.D, Statistics and Biostatistics

REGRESSION MODELS FOR DYNAMIC TREATMENT REGIMENS AND QUANTILE ASSOCIATION OF BIVARIATE SURVIVAL DATA

Ling-Wan Chen, PhD

University of Pittsburgh, 2018

In this dissertation we propose two new regression models under different types of survival data, including regression analysis for cumulative incidence functions (CIFs) under two-stage randomization, and quantile association regression for bivariate survival data.

The first topic concerns dynamic treatment regimens (DTRs) which are sets of rules for choosing effective treatments for individual patients based on their characteristics and intermediate responses, and have drawn considerable attention in the field of personalized medicine. Sequential Multiple Assignment Randomized Trial (SMART) design is often used to gather data on different DTRs. In this dissertation, we focus on finding personalized optimal DTRs from a two-stage SMART by regressing covariates on CIFs for competing risks outcomes. To our best knowledge no regression is readily available for analyzing competing risks outcome data from a SMART. Thus, we extend existing CIF regression models to handle covariate effects for DTRs. Asymptotic properties are established for our proposed estimators. We show the improvement provided by our proposed methods through simulation studies, and illustrate its practical utility through an analysis of a two-stage neuroblastoma study, where disease progression is subject to competing-risk censoring by death.

In the second project, we focus on local association in bivariate survival times, which is often of scientific importance. The local association measures capture the dynamic pattern of association over time, and it is desirable to quantify local association for different characteristics of the population. In this work, we adopt a novel quantile-based local association measure, which is free of marginal distributions, and propose a quantile association

regression model to allow covariate effects on the local association under the copula framework. Estimating equations for the quantile association coefficients are constructed via the relationship between this quantile-based measure and the copula model. To avoid estimating density functions in variance estimation, we extend the induced smoothing idea to our proposed estimators in obtaining the covariance matrix. The asymptotic properties for the resulting estimators are studied. The proposed estimators and inference procedure are evaluated through simulation, and applied to an age-related macular degeneration (AMD) dataset in studying risk factors on the association between AMD progression in two eyes.

Keywords: Bivariate survival data; Competing risks; Conditional association; Copula; Fine and Gray; Induced Smoothing; Inverse probability weighting; Odds Ratio; Quantiles regression; Scheike model; Sequential Multiple Assignment Randomized Trial.

TABLE OF CONTENTS

1.0 INTRODUCTION	1
2.0 CUMULATIVE INCIDENCE REGRESSION FOR DYNAMIC TREATMENT REGIMENS	4
2.1 Introduction	4
2.2 Method	6
2.2.1 Setting and data	6
2.2.2 The Fine and Gray Model with Fixed Weights	9
2.2.3 The Fine and Gray Model with Time-Varying Weights	11
2.2.4 An extension to subjects without second-stage randomization	12
2.2.5 Extensions of the Scheike Model	13
2.3 Asymptotic Properties	15
2.4 Simulation	18
2.5 Analysis of Children's Neuroblastoma Study	24
2.6 Discussion	30
3.0 QUANTILE ASSOCIATION MODEL FOR BIVARIATE SURVIVAL DATA	31
3.1 Introduction	31
3.2 Method	34
3.2.1 Bivariate Survival Data and Models	34
3.2.2 Estimating Equations	36
3.2.3 Asymptotic Properties	39
3.2.4 Covariance Estimations	41

3.3 Simulation	45
3.4 Data Analysis	50
3.5 Discussion	55
4.0 REMARKS AND FUTURE WORKS	60
APPENDIX A. THE PROOF OF THEOREM 1	62
APPENDIX B. THE PROOF OF THEOREM 2	68
APPENDIX C. THE ASYMPTOTIC PROPERTIES FOR THE FIXED WEIGHT SCHEIKE MODEL	72
APPENDIX D. THE PROOF OF THEOREM 3	77
APPENDIX E. THE PROOF OF THEOREM 4	80
APPENDIX F. JUSTIFICATION FOR THE CONSISTENCY OF THE PROPOSED INFLUENCE FUNCTION	86
BIBLIOGRAPHY	90

LIST OF TABLES

2.1	Simulation results for $A_1B_1B'_1$ and $A_1B_1B'_2$ with $n=400$ and $p_r=0.4$. The time point (time), the method for estimate (method), the true cumulative incidence (true), mean of estimates (est), empirical standard deviation ($\tilde{\sigma}$), mean of estimated standard deviations ($\hat{\sigma}$), coverage rate of 95% confidence intervals (Cov).	22
2.2	Simulation results for $A_1B_1B'_1$ and $A_1B_1B'_2$ with $n=400$ and $p_r=0.7$. The time point (time), the method for estimate (method), the true cumulative incidence (true), mean of estimates (est), empirical standard deviation ($\tilde{\sigma}$), mean of estimated standard deviations ($\hat{\sigma}$), coverage rate of 95% confidence intervals (Cov).	23
2.3	P-values for testing “time-varying effect” for each variable, using the Kolmogorov-Smirnov test and the Cramer-von Mises test.	28
2.4	The estimated coefficients from the WFG model for the four regimens. The estimate $\hat{\beta}$, the estimate of standard deviation ($\hat{\sigma}(\hat{\beta})$) and the P-value for testing $\beta = 0$	28
3.1	Simulation results for (I) marginal quantiles coefficients for the first event, (II) marginal quantiles coefficients for the second event, and (III) covariate effects on the quantile association under 20% censoring	48
3.2	Simulation results for (I) marginal quantiles coefficients for the first event, (II) marginal quantiles coefficients for the second event, and (III) covariate effects on the quantile association under 30% censoring	49

3.3	Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 20% censoring rate, n=400, and 2000 runs	51
3.4	Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 20% censoring rate, n=600, and 2000 runs	52
3.5	Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 50% censoring rate, n=400, and 2000 runs	53
3.6	Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 50% censoring rate, n=600, and 2000 runs	54

LIST OF FIGURES

2.1	A two-stage SMART setup	7
2.2	The estimated CIFs over time using six models. The black solid line is the true function. Gray lines for Fine-Gray-related models. Black stepwise curves for Scheike-related models. The native methods are dashed lines, the fixed weight methods are dotted lines and the time-varying weight are long dashed lines.	21
2.3	The estimated coefficient for regimen ABMT/cis-RA using the fixed weight Scheike model. The solid lines are estimates along with their confidence intervals (dashed lines) and confident bands (dotted lines).	27
2.4	The estimated CIFs for the four regimens obtained by using the WFG method with four cases while controlling for Age=3 years, Stage4dx=0 and Bonedx=0. The plots in the upper row are for Ferritindx=0 and the plots in the lower row are for Ferritindx=1. The plots in the left column are for MYCNdx=0 and the plots in the right column are for MYCNdx=1.	29
3.1	The estimated covariate effects for the left eye quantile. The solid bold line is the estimated effect at each quantile level and the dotted line is the average effect across all levels. The dash-dot line is the 95% pointwise confidence interval.	56
3.2	The estimated covariate effects for the right eye quantile. The solid bold line is the estimated effect at each quantile level and the dotted line is the average effect across all levels. The dash-dot line is the 95% pointwise confidence interval.	57

3.3	The estimated covariate effects on the association using the proposed model.	
	The solid bold line is the estimated effect at each quantile level and the dotted	
	line is the average across all levels. The dash-dot line is the 95% pointwise	
	confidence interval.	58

1.0 INTRODUCTION

This dissertation consists of two projects, which address methodological challenges in analyzing different types of survival data via the regression framework. In the first project, we focus on univariate competing risks data that are collected from a two-stage randomized trial. The other project is to establish an association regression model for bivariate survival data. The following two chapters detail methodological developments for these two specific topics.

In Chapter 2, we focus on finding the optimal dynamic treatment regimen (DTR). A DTR is a sequence of decision rules that one makes at each stage of intervention or treatment. An optimal DTR is to personalize treatments based on patients' past treatments, response status and key covariates, in order to achieve long-term optimal outcomes. When an event-type of outcome is of interest, the problem becomes identifying the best DTR for which patients have the smallest (or largest) probability of developing the target adverse (or beneficial) event. A sequential multiple assignment randomized trial (SMART) design has been proposed specifically for the purpose of developing optimal DTRs (Murphy, 2003). In this chapter, we focus on evaluating DTRs from a two-stage randomized trial, when the outcome of interest is subject to competing-risk censoring. As the competing events exist, researchers often use the cumulative incidence function (CIF) to quantify the cumulative risk of the target event by a specific time point. However, the standard nonparametric estimators of the CIF and the CIF regression models cannot be directly applied to SMART data. Recently, Yavuz et al. (2018) extended the nonparametric estimators for the CIFs to a two-stage randomization setting, without using the information on patients' characteristics. To personalize treatments based on patients' unique characteristics and history, it is crucial to incorporate covariates into the CIF estimation, which subsequently allows the optimal DTRs

to vary for heterogeneous patients. Therefore, we propose to extend some existing CIF regression models to the two-stage randomization setting. If the regression model is correctly specified, the resulting CIF estimators are shown to be consistent and can be approximated by a Gaussian process using empirical process theories. The finite-sample performance and the robustness of the CIF estimators under mis-specified models are evaluated through extensive simulations. We apply our proposed methods to a two-stage neuroblastoma study. Based on the estimated CIFs given covariates, an optimal DTR is recommended for patients with different risk factors, in order to minimize their risks of disease progression.

In Chapter 3, we focus on exploring the association between two event times. The association analyses are useful, since understanding how two events are related helps scientists to develop strategies to prevent or promote the occurrence of an event, when they observe the associated event. For bivariate survival data, several global association measures were proposed, such as Kendall’s tau and the correlation between two cumulative variates. However, they cannot capture the dynamic local association pattern over time. Various local association measures have been proposed via a copula framework (Oakes, 1989; Anderson et al., 1992; Shih and Louis, 1995; Nan et al., 2006; Hu and Nan, 2011), because the copula model allows time-dependent association between two failure times, and the estimation of the copula parameters is independent of marginal distributions. Furthermore, assessing the potential risk factors in the association analysis is of scientific importance, where the conditional association is adjusted for confounder effects and potential predictors. Many studies have been proposed to allow covariate effects on marginal distributions only (Zeng et al., 2009; Li et al., 2016). In practice, risk factors may affect the local association directly, in addition to their effects on the marginal distributions. To handle this challenge, we adopt a novel quantile-specific association measure as proposed in Li et al. (2014), which is independent of the marginal distributions, and establish a quantile association model to allow covariate effects on this quantile-based association measure between two failure times. However, Li et al. (2014) only dealt with completely observed bivariate data. Thus in this chapter, we develop an estimating equation for the quantile association coefficients via the relationship between this quantile-based measure and the copula. The asymptotic properties of the proposed estimator are established using the counting process approach under some

mild regularity conditions and the univariate censoring setting. The further challenge in this work is the covariance estimation in the analyses of quantile regression and quantile association due to non-smooth objective functions. To address this issue, we adapt the induced smoothing technique (Pang et al., 2012) to our quantile association analysis setting, and show that the estimated covariance is consistent. We apply the proposed estimators and the covariance procedure to numerical simulations and the data from an age-related macular degeneration (AMD) study to investigate the association of developing AMD in both eyes.

The rest of the dissertation is organized as following. Chapter 2 details various regression models for the CIF that are adapted to the SMART setting. Chapter 3 focuses on quantile association models for the association between two event times. We conclude this dissertation with some discussions of future directions in Chapter 4.

2.0 CUMULATIVE INCIDENCE REGRESSION FOR DYNAMIC TREATMENT REGIMENS

2.1 INTRODUCTION

Dynamic treatment regimens (DTRs) are sets of decision rules for choosing effective treatments for individual patients, based on their characteristics and intermediate responses. Often practitioners are interested in finding the optimal DTR that leads to the most desirable outcome in the end. An efficient randomization design is the Sequential Multiple Assignment Randomized Trial (SMART), where patients are randomly assigned to the initial treatments and then randomized to available treatments in subsequent stages, as they become eligible. In this chapter, we focus on competing risks data from a two-stage randomization design that was motivated by a neuroblastoma study. Children in this study were first randomized to two initial treatments, and those who responded to the initial treatment were further randomized to receive one of the two maintenance options. Meanwhile, the event of interest, disease progression, cannot be observed after death.

If there were no competing-risk events, existing nonparametric methods could have been used. They either modeled a mean restricted survival time for a treatment regimen by using the inverse probability weighting (IPW) method ([Lunceford et al., 2002](#); [Wahed and Tsiatis, 2006](#)), or generated various weighted Kaplan-Meier (KM) estimators ([Guo and Tsiatis, 2005](#); [Miyahara and Wahed, 2010](#)), or proposed pattern-mixture estimators of the survival function of a DTR ([Tang and Wahed, 2015b](#)). However, competing-risk events, such as death, commonly occur when subjects are exposed to multiple failures, and the event of interest cannot be experienced with the occurrence of competing events. In the competing-risk literature, the cumulative incidence function (CIF) from a specific event is

often of interest and widely used, because it is easily interpretable and is non-parametrically identifiable. In a SMART design with competing risks endpoints, the objective generally is to find a regimen which results in a reduced probability of occurrence of the target event. Recently, [Yavuz et al. \(2018\)](#) proposed four weighted nonparametric estimators of the CIF for a specific DTR without accounting for patient heterogeneity (covariates). Thus, the focus of this study is to model covariate effects on the CIFs of different DTRs.

The Cox regression model ([Cox, 1972](#)) and the accelerated failure model ([Wei et al., 1990](#)) are two popular methods of modeling covariate effects on survival. [Fine and Gray \(1999\)](#) extended Cox regression to competing risks data, and proposed a proportional hazards model for CIFs. [Klein and Andersen \(2005\)](#) developed a parametric regression model on pseudo values of the CIF. [Scheike et al. \(2008\)](#) proposed a direct binomial regression to model the time-varying effects of covariates on the CIF, which is more flexible than the fixed-effect Fine and Gray model. Recently [Gerds et al. \(2012\)](#) proposed a multinomial logistic model that handles multiple competing causes, providing another flexible alternative to the Fine and Gray model. However, these approaches are not readily applicable to a SMART study.

In SMART literature, [Murphy \(2003\)](#) proposed a backward searching algorithm to minimize the regret function at each step and find the best DTR at K steps, considering previous history and decisions. [Zhao et al. \(2009\)](#) used reinforcement learning and Q-learning to discover personal optimal therapies on cancer trials. [Henderson et al. \(2010\)](#) proposed the regret-regression to predict outcomes based on the estimated regression coefficients, and to use the resulting residuals for model diagnostics. [Goldberg and Kosorok \(2012\)](#) introduced a novel approach on a multistage-decision problem with censored data by using Q-learning. [Tang and Wahed \(2015a\)](#) proposed a fixed weight estimator for the cumulative hazard function in a two-stage design, under a proportional hazards assumption. However, none of the above methods can be used directly for competing risks outcomes.

Hence, we extend some existing regression models from the competing risks literature to SMARTs, particularly to two-stage randomization settings, adopting the IPW idea to account for the second-stage randomization. Our proposed methods perform an unbiased estimator for the CIF under the two-stage randomization design, while considering the covariate effects and the presence of the competing risk. In addition, no computational cost in

the estimation is a benefit for the research which includes a complex treatment strategy. The rest of this chapter is organized as follows. We introduce two regression models in Sections 2.2.2 and 2.2.3, and extend the methods to the situation where subjects may develop the event before the second randomization in Section 2.2.4. To relax the assumption of Fine and Gray’s model, we further apply our idea to Scheike’s model in Section 2.2.5. Asymptotic properties of the methods are discussed in Section 2.3. Results from finite-sample simulations are given in Section 2.4 and the analysis of the Neuroblastoma study is given in Section 2.5. Finally we conclude with some remarks in Section 2.6.

2.2 METHOD

2.2.1 Setting and data

We consider a two-stage SMART as depicted in Figure 2.1. Subjects are first randomly assigned to an initial treatment, either A_1 or A_2 . Subjects who respond to the initial treatment are randomly assigned to either treatment B_1 or B_2 , and non-responders are randomized to treatments B'_1 or B'_2 . This results in eight DTRs $A_m B_k B'_l$ with $m, k, l = 1, 2$, where subjects will start with the initial treatment A_m , and receive B_k if they respond to A_m , or B'_l , otherwise. Define T^R as the time to the intermediate response since the initial randomization. The response to the initial treatment, $R(= 1, 0)$, is often determined if the response time is shorter than a pre-specified time period (e.g., achieving remission within 6 months). Let Z_1 and Z_2 be the second treatment indicators for the responders and for non-responders to the initial treatment. The long-term outcome of interest is subject to competing-risk events. Let T be the time to the first failure from K competing causes since the first randomization, and let $\epsilon \in (1, \dots, K)$ be the corresponding cause of failure, where $\epsilon = 1$ denotes the event of interest. In practice, the first failure might happen before the subjects respond to their first treatment. If death is the primary outcome of interest, for example, patients may die before they manage to achieve remission. Thus, we use S to denote randomization status, where $S = 0$ for subjects who only have the first randomization, and $S = 1$ for subjects whose

response to the first treatment can be observed and who enter the second randomization.

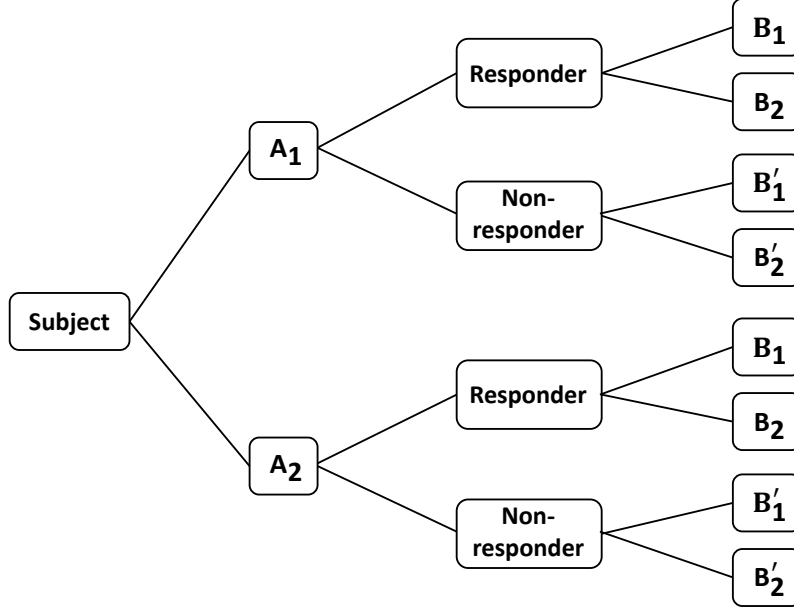


Figure 2.1: A two-stage SMART setup

Without any loss of generality, we focus on the regimens starting with the initial treatment A_1 . For a particular DTR, $A_1B_kB'_l, k, l = 1, 2$, we define the event time as $T_{A_1B_kB'_l}$ and the corresponding cause of failure as $\epsilon_{A_1B_kB'_l}$. Let T_{A_1} and ϵ_{A_1} be the event time and the cause indicator when a subject following $A_1B_kB'_l$ has developed the event of interest before the second randomization. If the subject proceeds to the second randomization, and is further randomized to B_k , we define the corresponding event time and the cause indicator as $T_{A_1B_k}$ and $\epsilon_{A_1B_k}$. $T_{A_1B'_l}$ and $\epsilon_{A_1B'_l}$ are similarly defined for the treatment path $A_1B'_l$. Thus, $T_{A_1B_kB'_l} = I(S = 0)T_{A_1} + I(S = 1, R = 1)T_{A_1B_k} + I(S = 1, R = 0)T_{A_1B'_l}$, and $\epsilon_{A_1B_kB'_l} = I(S = 0)\epsilon_{A_1} + I(S = 1, R = 1)\epsilon_{A_1B_k} + I(S = 1, R = 0)\epsilon_{A_1B'_l}$.

Note that T_{A_1} , ϵ_{A_1} , $T_{A_1B_k}$, $\epsilon_{A_1B_k}$, $T_{A_1B'_l}$, and $\epsilon_{A_1B'_l}$ are all counterfactuals, since a subject who is assigned to the DTR $A_1B_kB'_l$ can only follow one of the three potential paths. Here we adopt the consistency assumption (Hernan and Robins, 2010) in that if a subject follows a particular path, e.g., $S = 0$, the observed event time and the cause indicator for this subject are the same as the counterfactuals T_{A_1} and ϵ_{A_1} . Under the random assignment of treatments,

“no unmeasured confounders” and “positivity” assumptions are satisfied (Orellana et al., 2010). Here we consider a more general setting where some subjects have developed the event of interest before the second randomization, as in our neuroblastoma example. If none of the events occur during the first stage of randomization, T_{A_1} and ϵ_{A_1} , which are the counterfactuals for those subjects without proceeding to the second randomization, would become irrelevant and should be dropped off from the definition of $T_{A_1 B_k B'_l}$ and $\epsilon_{A_1 B_k B'_l}$ to ensure positivity. As we are considering SMART studies, exchangeability naturally follows as the probability of subsequent assignment is independent of potential outcomes given covariates and treatment history up to this point.

Let \mathbf{X} be a $p \times 1$ time-independent covariate vector. We are interested in evaluating covariate effects on the cause-1 CIF of a DTR $A_1 B_k B'_l$. That is,

$$F_{1,A_1 B_k B'_l}(t; \mathbf{X}) = \text{pr}(T_{A_1 B_k B'_l} \leq t, \epsilon_{A_1 B_k B'_l} = 1 \mid \mathbf{X}), k, l = 1, 2. \quad (2.1)$$

With the definition of counterfactuals, we then further define the i th patient’s event time as $T_i = (1 - S_i)T_{A_1 i} + S_i R_i \sum_{k=1}^2 I(Z_{1i} = k)T_{A_1 B_k i} + S_i(1 - R_i) \sum_{l=1}^2 I(Z_{2i} = l)T_{A_1 B'_l i}$, and the corresponding cause of failure as $\epsilon_i = (1 - S_i)\epsilon_{A_1 i} + S_i R_i \sum_{k=1}^2 I(Z_{1i} = k)\epsilon_{A_1 B_k i} + S_i(1 - R_i) \sum_{l=1}^2 I(Z_{2i} = l)\epsilon_{A_1 B'_l i}$. Since we adopt the consistency assumption (Rubin, 1974) to relate the uncensored survival time T_i to the counterfactual outcomes, for the i subject who is assigned to Regimen $A_1 B_k B_l$, we have that the observed uncensored outcome is equal to the corresponding counterfactual outcome, i.e. $T_i = T_{A_1 B_k B_l i}$, for $k, l = 1, 2$. Under the random assignment of treatments, the “no unmeasured confounders” and “positivity” assumptions are satisfied in this counterfactual model. In general, there may be right censoring C before we observe T , and we assume that C and T are conditionally independent given baseline covariates. Let C be the potential censoring time with $G(t) = \text{pr}(C > t)$. In the presence of conditionally independent censoring, one observes $V = \min(T, C)$, $\Delta = I(T \leq C)$ and ϵ . Competing risks data from a two-stage SMART trial consist of n independent and identically distributed copies of $\{S_i, S_i R_i, S_i R_i Z_{1i}, S_i(1 - R_i)Z_{2i}, V_i, \Delta_i, \Delta_i \epsilon_i, \mathbf{X}_i\}_{i=1}^n$.

2.2.2 The Fine and Gray Model with Fixed Weights

Fine and Gray (1999) proposed a semiparametric proportional hazards model for the subdistribution of a competing risk, and assumed that

$$g\{F_1(t; X)\} = h_0(t) + \mathbf{X}^T \beta_0,$$

where $g(u) = \log(-\log(1 - u))$, and $h_0(\cdot)$ is a completely unspecified, invertible, and monotone increasing function. If we define the hazard function for the CIF (or subdistribution) $\lambda_1(t) = d \log F_1(t)/dt$, the above model has the proportional hazards interpretation for the subdistribution hazards, where $\lambda_1(t; \mathbf{X}) = \lambda_{10}(t) \exp(\mathbf{X}^T \beta_0)$, with $\lambda_{10}(t)$ being the baseline hazard function at time t . For a particular DTR $A_1 B_k B'_l$ in Fine and Gray's model, the CIF in (2.1) can be formulated as

$$\mathbf{F}_{1, A_1 B_k B'_l}(t; \mathbf{X}) = 1 - \exp \left\{ - \int_0^t \lambda_{10, A_1 B_k B'_l}(u) \exp(\mathbf{X}^T \beta_{0, A_1 B_k B'_l}) du \right\}, k, l = 1, 2. \quad (2.2)$$

where $\lambda_{10, A_1 B_k B'_l}(t)$ is the baseline subdistribution hazard function at time t in $A_1 B_k B'_l$ DTR, and $\beta_{0, A_1 B_k B'_l}$ is the coefficient vector in $A_1 B_k B'_l$ DTR. To simplify the notation, we use $\lambda_{10}(t)$ to denote $\lambda_{10, A_1 B_k B'_l}(t)$ and refer to $\beta_{0, A_1 B_k B'_l}$ as β_0 in the rest of paper, if there is no confusion. For data from two-stage randomized trials, if we apply the Fine and Gray method directly to estimate the CIF for $A_1 B_k B'_l$, only the data from subjects following treatment sequences $A_1 B_k$ or $A_1 B'_l$ are included in the estimation of (2.1). The estimated CIF is often biased, since this naive Fine and Gray method weighs each subject consistent with $A_1 B_k B'_l$ equally in the estimation. To see the potential bias, let us consider a hypothetical example. Suppose 100 subjects are randomized to follow the DTR $A_1 B_k B'_l$, and 40 subjects respond to the initial treatment A_1 and the rest do not. If there were no second-stage randomization, we would expect the 40 responders to follow the treatment sequence $A_1 B_k$ and the 60 non-responders to follow $A_1 B'_l$. An unbiased estimate of the CIF for $A_1 B_k B'_l$ will include the information from these 40 responders and 60 non-responders. Now with the second randomization, suppose 20 responders are assigned to the second-stage treatment B_k and 18 non-responders are assigned to B'_l . The sub-sample used by the naive Fine-Gray method consists of these 20 responders and 18 non-responders, which has a higher proportion of responders as compared to the

original sample without the second-stage randomization. Therefore, if we treat responders and non-responders in this subsample equally, we tend to have a biased estimate of the CIF.

To account for the bias, we follow a similar IPW approach as in [Guo and Tsiatis \(2005\)](#), [Miyahara and Wahed \(2010\)](#), and [Yavuz et al. \(2018\)](#). Since the proportion of responders in the sub-sample, which is used in the naive Fine and Gray method, is not the same as that in the original sample before second-stage randomization, we assign the responders and non-responders in the sub-sample weights that are inversely proportional to their probabilities of being assigned to B_k or B'_l . In the created pseudo sub-sample, the sizes of responders and non-responders are about the same as the original sample. More specifically, let π_{B_k} and $\pi_{B'_l}$ be the true probabilities of being assigned to B_k for responders and being assigned to B'_l for non-responders, where $\pi_{B_k} = \text{pr}(Z_{1i} = k \mid R_i = 1)$ and $\pi_{B'_l} = \text{pr}(Z_{2i} = l \mid R_i = 0)$. Define

$$Q_{A_1 B_k B'_l, i} = R_i I(Z_{1i} = k) / \pi_{B_k} + (1 - R_i) I(Z_{2i} = l) / \pi_{B'_l}$$

as the weight for subject i in the $A_1 B_k B'_l$ regimen. Due to randomization, the observed proportions of being assigned to sequences $A_1 B_k$ and $A_1 B'_l$ are not exactly equal to the true probabilities, and consequently, they may provide more information about the randomization process. Thus, we consider using the estimated probabilities, $\hat{\pi}_{B_k}$ and $\hat{\pi}_{B'_l}$, from the sequences $A_1 B_k$ and $A_1 B'_l$, instead of the true probabilities, to obtain the estimated fixed weight, $\hat{Q}_{A_1 B_k B'_l, i} = R_i I(Z_{1i} = k) / \hat{\pi}_{B_k} + (1 - R_i) I(Z_{2i} = l) / \hat{\pi}_{B'_l}$, for subject i . As a result, the pseudo sample that this estimated fixed weight $\hat{Q}_{A_1 B_k B'_l}$ creates, has the exactly same number of subjects and the same mixture of responders and non-responders as the original sample.

Here we extend Fine and Gray's model to the two-stage randomized trials with estimated fixed weights. For subject i , similar to [Fine and Gray \(1999\)](#), we defined the weight as $w_i(t) = I(C_i \geq T_i \wedge t) \hat{G}(t) / \hat{G}(V_i \wedge t)$, where $\hat{G}(\cdot)$ is the Kaplan-Meier estimate of the survival function for censoring C . To avoid the same potential bias of using the sub-sample after second randomization, we construct \hat{G} based on the weighted counting and at-risk processes $N_{1i}^w(t) = I(V_i \leq t, \epsilon_i = 1) \hat{Q}_{A_1 B_k B'_l, i}$ and $Y_i^{*w}(t) = \{1 - I(V_i \leq t, \epsilon_i = 1)\} \hat{Q}_{A_1 B_k B'_l, i}$. Combining their weighted vital status to handle censored observations and our estimated fixed weights

to let the sub-sample represent all related responders and non-responders, we define the fixed weight score function for the $A_1 B_k B'_l$ regimen as

$$U_{A_1 B_k B'_l}(\beta) = \sum_{i=1}^n \int_0^\infty \left\{ \mathbf{X}_i - \frac{\sum_{j=1}^n w_j(u) Y_j^{*w}(u) \mathbf{X}_j \exp(\mathbf{X}_j^T \beta)}{\sum_{j=1}^n w_j(u) Y_j^{*w}(u) \exp(\mathbf{X}_j^T \beta)} \right\} w_i(u) dN_{1i}^w(u). \quad (2.3)$$

Let $\hat{\beta}$ be a solution to the above score equation (2.3). In our study, the estimator of the CIF is of greater interest. Based on the estimated $\hat{\beta}$, we evaluate the CIF at time t_0 with covariates \mathbf{x}_0 by using the formula in (2.2), namely,

$$\hat{F}_{1, A_1 B_k B'_l}(t_0; \mathbf{x}_0) = 1 - \exp\{-\hat{\Lambda}_1(t_0; \mathbf{x}_0)\},$$

where $\hat{\Lambda}_1(t_0; x_0)$ is the cumulative subdistribution hazard function at time t_0 with covariates x_0 estimated as

$$\hat{\Lambda}_1(t_0; x_0) = n^{-1} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \hat{\beta})}{n^{-1} \sum_{j=1}^n w_j(u) Y_j^{*w}(u) \exp(\mathbf{X}_j^T \hat{\beta})} w_i(u) dN_{1i}^w(u).$$

The limiting distribution of estimators and the inference are shown in Section 2.3.

2.2.3 The Fine and Gray Model with Time-Varying Weights

The weighted Fine and Gray method described in Section 2.2.2 does not utilize the information on time to response. Considering that subjects are consistent with all of the regimens before they have the intermediate response, we modify the above model using time-varying weights that incorporate those subjects with weights of 1 until their response status is observed. After obtaining their response status, subjects receive the weights according to their second randomization as in Section 2.2.2. More specifically, the weight for subject i at time t is

$$\hat{Q}_{A_1 B_k B'_l, i}(t) = \begin{cases} 1, & \text{if } T_i^R > t \\ \frac{R_i I(Z_{1i}=k)}{\hat{\pi}_{B_k}} + \frac{(1-R_i) I(Z_{2i}=l)}{\hat{\pi}_{B'_l}}, & \text{if } T_i^R \leq t. \end{cases}$$

This idea has been used in [Guo and Tsiatis \(2005\)](#), [Miyahara and Wahed \(2010\)](#) and [Yavuz et al. \(2018\)](#) for the nonparametric setting. The corresponding time-varying weight score function for the $A_1B_kB'_l$ regimen under Fine and Gray's model is

$$U_{A_1B_kB'_l}^{tw}(\beta) = \sum_{i=1}^n \int_0^\infty \left\{ \mathbf{X}_i - \frac{\sum_{j=1}^n w_j^{tw}(u) Y_j^{*tw}(u) \mathbf{X}_j \exp(\mathbf{X}_j^T \beta)}{\sum_{j=1}^n w_j^{tw}(u) Y_j^{*tw}(u) \exp(\mathbf{X}_j^T \beta)} \right\} w_i^{tw}(u) dN_{1i}^{tw}(u), \quad (2.4)$$

where $w_i^{tw}(u) = \{\hat{G}^{tw}(u)/\hat{G}^{tw}(V_i \wedge u)\}I(\Delta_i = 1) + I(V_i > u)$, $Y_i^{*tw}(t) = \{1 - I(V_i \leq t, \epsilon_i = 1)\}\hat{Q}_{A_1B_kB'_l,i}(t)$, and $N_{1i}^{tw}(t) = I(V_i \leq t, \epsilon_i = 1)\hat{Q}_{A_1B_kB'_l,i}(t)$, with \hat{G}^{tw} using the time-varying weighted counting and at-risk processes. Again the solution for (2.4), $\hat{\beta}^{tw}$, can be obtained via the Newton-Raphson algorithm. Hence, the time-varying weight estimator of the CIF, based on the estimated $\hat{\beta}^{tw}$, at time t_0 with covariates \mathbf{x}_0 is

$$\hat{F}_{1,A_1B_kB'_l}^{tw}(t_0; \mathbf{x}_0) = 1 - \exp\{-\hat{\Lambda}_1^{tw}(t_0; \mathbf{x}_0)\},$$

where

$$\hat{\Lambda}_1^{tw}(t_0; \mathbf{x}_0) = n^{-1} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(\mathbf{x}_0^T \hat{\beta}^{tw})}{n^{-1} \sum_{j=1}^n w_j^{tw}(u) Y_j^{*tw}(u) \exp(\mathbf{X}_j^T \hat{\beta}^{tw})} w_i^{tw}(u) dN_{1i}^{tw}(u).$$

The asymptotic properties of estimators and the inference are discussed in [Section 2.3](#).

2.2.4 An extension to subjects without second-stage randomization

In [Sections 2.2.2](#) and [2.2.3](#), our discussions have focused on the situations that all subjects enter the second randomization. In practice, it is likely that some subjects develop the event of interest before they respond to the initial treatment. These subjects are excluded from the analyses in the above models. However, if we assign all subjects to a specific regimen, e.g., $A_1B_kB'_l$, those subjects who have developed the event of interest before the second-stage randomization are following this regimen. Hence we expand our definition of “consistency” with the regimen, and now treat those subjects who have developed the target event before they meet the response criteria as consistent with the regimen. Consequently, we extend our methods by redefining the weights. If a subject has developed an event before their response status is determined, we assign the weight as 1 and record the event time. Such a subject

is denoted as $S = 0$ in Section 2.2.1, and the extended time-varying weight for subject i at time t is written as

$$\hat{Q}_{A_1 B_k B'_l, i}^*(t) = \begin{cases} 1, & \text{if } T_i^R > t \text{ and } S = 1; \text{ or } S = 0 \\ \frac{R_i I(Z_{1i}=k)}{\hat{\pi}_{B_k}} + \frac{(1-R_i) I(Z_{2i}=l)}{\hat{\pi}_{B'_l}}, & \text{if } T_i^R \leq t \text{ and } S = 1. \end{cases}$$

The score function in this case is given by replacing $\hat{Q}_{A_1 B_k B'_l, i}(t)$ with $\hat{Q}_{A_1 B_k B'_l, i}^*(t)$ in (2.4). Estimation and inferences are very similar and hence the details are left out for brevity.

2.2.5 Extensions of the Scheike Model

Fine and Gray's model is popular and convenient in practice with the available software. If the proportionality for sub-distribution is satisfied, the results are accurate and easy to interpret. However, this assumption may be too restrictive for a two-stage randomization study, because the covariate effects on the CIF may change when subjects switch from the initial treatment to the second-stage treatment. Though the weighted Fine and Gray model in Sections 2.2.3 and 2.2.4 can still provide reasonable estimates of the CIF as shown in our simulation studies, we now consider extending a more flexible binomial regression model proposed by Scheike et al. (2008) to the two-stage randomization setting in order to capture potential time-varying covariate effects for a particular DTR. The additive Scheike model assumes that

$$F_1^{\eta, \gamma}(t, \mathbf{X}) = h\{\mathbf{X}_1^T \eta(t) + g(\gamma, \mathbf{X}_2, t)\},$$

where $\eta(t)$ are the time-varying effects of \mathbf{X}_1 , a subset of covariates, on the CIF at time t , and γ are the fixed-effect coefficients for the rest of covariates, \mathbf{X}_2 . The h and g are known link functions. If $h(x_1) = 1 - \exp(-x_1)$ and $g(\gamma, x_2) = \exp(\gamma^T x_2)$, the Scheike model will become the proportional hazards model for subdistributions as in Fine and Gray (1999).

As before, we extend the original Scheike model to the fixed weight Scheike model for a two-stage randomization setting. Let $F_{1, A_1 B_k B'_l, i}^{\eta, \gamma}(t; \mathbf{X}_i)$ denote the cause-1 CIF at time t for subject i with covariates \mathbf{X}_i following the regimen $A_1 B_k B'_l$. The estimating equation for $A_1 B_k B'_l$ at time t can be written as $U^*(\eta, \gamma, \hat{G}) = \{U_1^*(\eta, \gamma, \hat{G})(t), U_2^*(\eta, \gamma, \hat{G})\}$, where

$$U_1^*(\eta, \gamma, \hat{G})(t) = \sum_{i=1}^n D_{\eta, i}^{*T} u_i(t) \left\{ \frac{\Delta_i N_{1i}(t)}{\hat{G}_i(V_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta, \gamma) \right\} \hat{Q}_{A_1 B_k B'_l, i},$$

$$U_2^*(\eta, \gamma, \hat{G}) = \sum_{i=1}^n \int_a^\tau D_{\gamma,i}^{*T} u_i(t) \left\{ \frac{\Delta_i N_{1i}(t)}{\hat{G}_i(V_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta, \gamma) \right\} \hat{Q}_{A_1 B_k B'_l, i} dt,$$

with $F_{1i}^*(t; \eta, \gamma)$ denoting $F_{1, A_1 B_k B'_l, i}^{\eta, \gamma}(t; \mathbf{X}_i)$ for brevity, $D_{\eta, i}^{*T} = \partial F_{1i}^*(t; \eta, \gamma) / \partial \eta(t)$, $D_{\gamma, i}^{*T} = \partial F_{1i}^*(t; \eta, \gamma) / \partial \gamma$ denoting partial derivatives, and $u_i(t)$ being some possibly random weights.

We use binomial regression as in [Scheike et al. \(2008\)](#), coupled with the Newton-Raphson iteration, to obtain the estimator $\hat{\eta}(t)$ for the time-varying coefficients at each time point t , and the estimator $\hat{\gamma}$ for the time-independent coefficients. The estimation of the CIF for given covariates can be carried out similar to the extended Fine and Gray models. Furthermore, we can establish the Scheike model with time-varying weights by using the estimated time-varying weights in Section 2.2.3 ($\hat{Q}_{A_1 B_k B'_l}(t)$), and also extend this model so that subjects without the second stage randomization are included by using the weights defined in Section 2.2.4 ($\hat{Q}_{A_1 B_k B'_l}^*(t)$).

However, the inferences of these estimators are much more involved. In Appendix C, we give the influence functions for $\hat{\eta}$ and $\hat{\gamma}$ under the simplified setting where all events occur only after second-randomization. We implement the fixed-weighted Scheike model and the time-varying weight Scheike model by treating all covariates with time-varying coefficients. The implementation is rather complicated. Therefore, we also propose an approximation based on the idea of augmenting the data for the fixed weight Scheike model. To illustrate our idea, we continue to consider the hypothetical example for $A_1 B_k B'_l$ in Section 2.2.2, where 100 subjects are assigned to the initial treatment A_1 , and 40 of them respond to A_1 . During the second-stage randomization, 20 of the 40 responders are assigned to B_k , and 18 of the 60 nonresponders are assigned to B'_l . According to Section 2.2.2, the fixed weight for subjects in the sequence of $A_1 B_1$ is 2, and that for subjects following $A_1 B'_1$ is 3.33 $\bar{3}$. We create an augmented data by repeating each subject in $A_1 B_1$ 20 times, and each subject following $A_1 B'_1$ 33 times. This augmented data contains 400 responders and 594 non-responders, approximately the same mixture of responders and non-responders as the original sample. Thus, a well-implemented R function, “comp.risk”, in the package **timereg** for the [Scheike et al. \(2008\)](#), can be directly applied to the augmented data, resulting in a consistent, though slightly less accurate, estimator of the CIF for $A_1 B_1 B'_1$. Since the size of the augmented data is about 10 times of the original sample, the standard deviation from

the R function needs to be multiplied by the squared root of the ratio of the augmented data sample size to the original sample size. In general, the augmented data approach can be applied to other models, such as the fixed weight Fine and Gray method.

2.3 ASYMPTOTIC PROPERTIES

In this section, we establish asymptotic properties of our proposed estimators. Because the fixed weight Fine and Gray method is a special case of the time-varying weight Fine and Gray model, we focus on time-varying weight Fine and Gray model in this section.

In the inference on the estimation of CIFs, we use the weights included the true probabilities, π_{B_k} and $\pi_{B'_l}$. Let $Q_{A_1 B_k B'_l, i}(t)$ be 1 if $T_i^R < t$ and be $Q_{A_1 B_k B'_l, i}$, otherwise. Define $\tilde{N}_{1i}^{tw}(t) = Q_{A_1 B_k B'_l, i}(t) N_{1i}(t)$ as a counting process and $\tilde{M}_{1i}^{tw}(t, \beta_0) = \tilde{N}_{1i}^{tw}(t) - \tilde{A}_i^{tw}(t, \beta_0)$ as a martingale, where $\tilde{A}_i^{tw}(t, \beta) = \int_0^t \tilde{Y}_i^{*tw}(s) \lambda_{10}(s) \exp(\mathbf{X}_i^T \beta) ds$ and $\tilde{Y}_i^{*tw}(s) = \{1 - N_{1i}(s-)\} Q_{A_1 B_k B'_l, i}(s)$. We replace $\hat{Q}_{A_1 B_k B'_l, i}(t)$ by $Q_{A_1 B_k B'_l, i}(t)$ in the time-varying weighted score function, and recast this score function in terms of martingale integration, under the true β_0 . To simplify the score equation, denote $S^{tw(p)}(\beta, u) = n^{-1} \sum_{i=1}^n w_i^{tw}(u) \tilde{Y}_i^{*tw}(u) \mathbf{X}_i^{\otimes p} \exp(\mathbf{X}_i^T \beta)$, $p = 0, 1, 2$, and $\bar{\mathbf{X}}^{tw}(\beta, u) = S^{tw(1)}(\beta, u) / S^{tw(0)}(\beta, u)$, where $S^{tw(1)}(\beta, u) = \partial S^{tw(0)}(\beta, u) / \partial \beta$ and $S^{tw(2)}(\beta, u) = \partial^2 S^{tw(0)}(\beta, u) / \partial \beta \partial \beta^T$. The time-varying weight score function for the $A_1 B_1 B'_1$ regimen, under the true β_0 , is

$$U_{A_1 B_k B'_l}^{tw}(\beta_0) = \sum_{i=1}^n \int_0^\infty \{\mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta_0, u)\} w_i^{tw}(u) d\tilde{M}_{1i}^{tw}(u, \beta_0).$$

Theorem 1. *Under mild regularity conditions, the $n^{-1/2} U_{A_1 B_k B'_l}^{tw}(\beta_0)$ converges in distribution to a Gaussian process with covariance matrix Σ^{tw} . Then, the asymptotical distribution of $n^{1/2}(\hat{\beta}^{tw} - \beta_0)$ is normally distributed with mean zero and the covariance matrix $\Omega^{tw-1} \Sigma^{tw} \Omega^{tw-1}$, where*

$$\Omega^{tw} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^\infty \left\{ \frac{S^{(2)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} - \bar{\mathbf{X}}(\beta_0, u)^{\otimes 2} \right\} w_i(u) d\tilde{N}_{1i}^{tw}(u).$$

The proof of Theorem 1 and the form of Σ^{tw} are included in Appendix A. Consistent estimators of Ω^{tw} and Σ^{tw} are

$$\hat{\Omega}^{tw} = \frac{1}{n} \sum_{i=1}^n Q_{A_1 B_k B'_l, i}(V_i) \left\{ \frac{S^{tw(2)}(\hat{\beta}^{tw}, V_i)}{S^{tw(0)}(\hat{\beta}^{tw}, V_i)} - \bar{\mathbf{X}}^{tw}(\hat{\beta}^{tw}, V_i)^{\otimes 2} \right\} \Delta_i I(\epsilon_i = 1), \text{ and}$$

$$\hat{\Sigma}^{tw} = n^{-1} \sum_{i=1}^n \left(\hat{\eta}_i^{tw} + \hat{\psi}_i^{tw} \right)^{\otimes 2},$$

where

$$\hat{\eta}_i^{tw} = \int_0^\infty \left\{ \mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\hat{\beta}^{tw}, u) \right\} w_i^{tw}(u) d\hat{M}_{1i}^{tw}(u, \hat{\beta}^{tw}), \text{ and}$$

$$\hat{\psi}_i^{tw} = \int_0^\infty \frac{\hat{q}^{tw}(s, \hat{\beta}^{tw})}{\hat{\pi}^{tw}(s)} d\hat{M}_i^{c, tw}(s).$$

We use $Q_{A_1 B_k B'_l, i}(u)$ in computing the weighted vital status, $w_i^{tw}(u) = \{\tilde{G}^{tw}(u)/\tilde{G}^{tw}(V_i \wedge u)\} I(\Delta_i = 1) + I(V_i > u)$, where $\tilde{G}^{tw}(u)$ is the Kaplan-Meier estimator with the sub-sample re-represented using $Q_{A_1 B_k B'_l, i}(u)$. More specifically,

$$\tilde{G}^{tw}(u) = \prod_{V_j \leq u} \left[1 - \left\{ \sum_{i=1}^n Q_{A_1 B_k B'_l, i}(u) I(V_i = u, \Delta_i = 0) \right\} / \left\{ \sum_{i=1}^n Q_{A_1 B_k B'_l, i}(u) I(V_i \geq u) \right\} \right].$$

Also,

$$\begin{aligned} \hat{M}_{1i}^{tw}(u, \hat{\beta}^{tw}) &= Q_{A_1 B_k B'_l, i}(u) I(V_i \leq u, \epsilon_i = 1) \\ &\quad - \int_0^u Q_{A_1 B_k B'_l, i}(t) \{1 - I(V_i < t, \epsilon_i = 1)\} \exp(\mathbf{X}_i^T \hat{\beta}^{tw}) d\hat{\Lambda}_{10}^{tw}(t) \end{aligned}$$

is the estimated Martingale for the cause-1 event, where

$$\hat{\Lambda}_{10}^{tw}(t) = n^{-1} \sum_{i=1}^n \int_0^t \{w_i^{tw}(u)/S^{tw(0)}(\hat{\beta}^{tw}, u)\} d\tilde{N}_{1i}^{tw}(u).$$

In $\hat{\psi}_i^{tw}$, $\hat{q}^{tw}(s, \hat{\beta}^{tw}) = -n^{-1} \sum_{j=1}^n \int_0^\infty [\mathbf{X}_j - \bar{\mathbf{X}}^{tw}(\hat{\beta}^{tw}, u)] w_j^{tw}(u) d\hat{M}_{1j}^{tw}(u, \hat{\beta}^{tw}) I(V_j < s \leq u)$, $\hat{\pi}^{tw}(s) = n^{-1} \sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)$, and $\hat{M}_i^{c, tw}(s) = Q_{A_1 B_k B'_l, i}(s) I(V_i \leq s, \epsilon_i = 0) - \int_0^s Q_{A_1 B_k B'_l, i}(t) I(V_i \geq t) d\hat{\Lambda}^{c, tw}(t)$ is the estimated martingale for censoring, where $\hat{\Lambda}^{c, tw}(t) = \int_0^t \sum_{i=1}^n Q_{A_1 B_k B'_l, i}(u) / \left\{ \sum_{j=1}^n Q_{A_1 B_k B'_l, j}(u) I(V_i \geq u) \right\} dI(V_i = u, \Delta_i = 0)$.

Theorem 2. *If a consistent estimator $\hat{\beta}^{tw}$ exists and $n^{1/2}(\hat{\Lambda}_1^{tw}(t_0, \mathbf{x}_0) - \Lambda_1(t_0, \mathbf{x}_0))$ converges in distribution to a Gaussian process on an interval $[0, c)$, where $\text{pr}(X \geq c) > 0$, then $n^{1/2}\{\hat{F}_{1,A_1B_kB'_l}^{tw}(t_0; \mathbf{x}_0) - F_{1,A_1B_kB'_l}(t_0; \mathbf{x}_0)\}$ has the same limiting distribution as*

$$K^{tw}(t_0; \mathbf{x}_0) = n^{-1/2} \exp\{-\Lambda_1(t_0, \mathbf{x}_0)\} \times \left\{ \sum_{i=1}^n \int_0^{t_0} \frac{\exp(\mathbf{x}_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta_0) \right. \\ \left. + \sum_{j=1}^n \int_0^\infty \frac{v^{tw}(s, t_0, \mathbf{x}_0, \beta_0)}{\pi^{tw}(s)} d\tilde{M}_j^{c,tw}(s) + \mathbf{h}^{twT}(t_0, \mathbf{x}_0) \Omega^{tw-1} \sum_{i=1}^n (\eta_i^{tw} + \psi_i^{tw}) \right\} + o_p(1), \quad (2.5)$$

where

$$v^{tw}(s, t_0, \mathbf{x}_0, \beta_0) = - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(\mathbf{x}_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \tilde{w}_i(u) I(V_i < s \leq u) d\tilde{M}_{1i}^{tw}(u, \beta_0),$$

$$\pi^{tw}(s) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n Q_{A_1B_kB'_l,m}(s) I(V_m \geq s),$$

and

$$\mathbf{h}^{tw}(t_0, \mathbf{x}_0) = \int_0^{t_0} \{\mathbf{x}_0 - \bar{\mathbf{x}}^{tw}(\beta_0, u)\} \exp(\mathbf{x}_0^T \beta_0) d\Lambda_{10}(u).$$

The detailed proofs of Theorem 2 are given in Appendix B. However, it is complicated to evaluate the exact limiting distribution of the CIF estimator in (2.5). Hence we follow Fine and Gray (1999) and adopt an approximation based on random perturbation. More specifically, let $\{\mathbf{W}_i\}_{i=1}^n$ be a random sample from the standard normal distribution and

$$\hat{K}^{tw}(t_0, x_0) = n^{-1/2} \exp\{-\hat{\Lambda}^{tw}(t_0, \mathbf{x}_0)\} \times \left\{ \sum_{i=1}^n \int_0^{t_0} \frac{\exp(\mathbf{x}_0^T \hat{\beta}^{tw})}{\hat{S}^{tw(0)}(\hat{\beta}^{tw}, u)} w_i^{tw}(u) d\hat{M}_{1i}^{tw}(u, \hat{\beta}^{tw}) \mathbf{W}_i \right. \\ \left. + \sum_{i=1}^n \int_0^\infty \frac{\hat{v}^{tw}(s, t_0, \mathbf{x}_0, \hat{\beta}^{tw})}{\hat{\pi}^{tw}(s)} d\hat{M}_i^{c,tw}(s, \hat{\beta}^{tw}) \mathbf{W}_i + \hat{\mathbf{h}}^{twT}(t_0, \mathbf{x}_0) \hat{\Omega}^{tw-1} \sum_{i=1}^n (\hat{\eta}_i^{tw} + \hat{\psi}_i^{tw}) \mathbf{W}_i \right\}.$$

To obtain the estimated variance at t_0 with covariates \mathbf{x}_0 , we generate B samples $\{\mathbf{W}_{bi}, i = 1, \dots, n\}$, $b = 1, \dots, B$, and compute $\hat{K}_b^{tw}(t_0, \mathbf{x}_0)$ for $b = 1, \dots, B$. Then the standard deviation for the CIF estimator at time t_0 can be estimated by

$$\hat{\sigma}^{tw}(t_0, \mathbf{x}_0) = \left\{ (nB)^{-1} \sum_{b=1}^B \hat{K}_b^{tw^2}(t_0, \mathbf{x}_0) \right\}^{1/2}.$$

We further discuss the extension of models with $\hat{Q}_{A_1B_kB'_l,j}(t)$. Several studies have shown that the inference results remain similar as before, even when $Q_{A_1B_kB'_l,j}(u)$ is replaced by a

consistent estimator (Yavuz et al., 2018, e.g.). Thus, we simply replace $Q_{A_1 B_k B'_l, j}(t)$ with $\hat{Q}_{A_1 B_k B'_l, j}(t)$ in the estimation of the CIFs and the inference procedures. All the above discussions can be applied to the fixed weight Fine and Gray model.

2.4 SIMULATION

We conduct extensive simulations to study the finite-sample performance of our proposed methods. Subjects are assumed to be randomized to the initial treatments A_1 and A_2 . Thus, we focus on only subjects who are assigned to A_1 . The following three scenarios are considered to mimic a general setting: 1) subjects have developed the event of interest before they respond to A_1 ; 2) subjects who have responded to A_1 within a specific time, say, 3 months (0.25 year), and have not developed either the target event or the competing event, are further randomized to treatments B_1 or B_2 ; and 3) subjects who have not responded to A_1 and have not developed any event within 3 months, are randomized to treatments B'_1 or B'_2 . Two covariates $\mathbf{X} = (X_1, X_2)$ are considered, where X_1 is a standard normal variate and X_2 is from Bernoulli(0.5). To create the three scenarios, we introduce T_1 as the time in years to either the first event (a cause-1 or competing event) or the intermediate response from the initial randomization. The subjects whose T_1 exceed 0.25 are treated as non-responders to the initial treatment A_1 . They are further randomized to B'_1 or B'_2 following a Bernoulli distribution with $p_2 = \text{pr}(Z_2 = 1 \mid S = 1, R = 0)$, and $S = 1$ and their response times T_1 are truncated at 0.25. Assume that T_1 is independent of the covariates and follows an Exponential distribution with rate λ . Here λ is determined by $p_s = \text{pr}(S = 0)$ and $p_r = \text{pr}(R = 1 \mid S = 1)$. Since $\text{pr}(T_1 \geq 0.25 \mid S = 1) = 1 - p_r$, then $\lambda = -\log\{(1 - p_r)(1 - p_s)\}/0.25$. For those with $T_1 < 0.25$, they can either have developed an event or responded before 0.25, whichever occurring first. Given $T_1 < 0.25$, we simulated $S = 0, 1$ following a Bernoulli distribution, where $\text{pr}(S = 0 \mid T_1 < 0.25) = p_s / \{1 - (1 - p_r)(1 - p_s)\}$. When $S = 0$, subjects are assumed to have developed an event before the second randomization, where the event cause indicator ϵ_1 is further simulated from a Bernoulli distribution with $\text{pr}(\epsilon_1 = 1 \mid S = 0) = 0.75$. When $S = 1$, those subjects are

deemed as responders with $R = 1$, and further randomized to B_1 or B_2 following a Bernoulli distribution with $p_1 = \text{pr}(Z_1 = 1 \mid S = 1, R = 1)$. In our simulations, we let $p_s = 0.1$, $p_r = 0.4$ or 0.7 , and $p_1 = p_2 = 0.3$.

For those subjects proceeding to the second randomization, we let T^{seq} and ϵ^{seq} denote the time to the first event and the corresponding cause indicator in a specific treatment sequence since the second randomization, where $seq = A_1B_1, A_1B_2, A_1B'_1$ and $A_1B'_2$. Assume that T^{seq} and ϵ^{seq} follow the Fine and Gray Model:

$$\text{pr}(T^{seq} \leq t, \epsilon^{seq} = 1 \mid S = 1, R, Z_1, Z_2, \mathbf{X}) = 1 - \{1 - p(1 - e^{-t})\}^{e^{\gamma_1 X_1 + \gamma_2 X_2}}, \quad (2.6)$$

where $p = 0.4$, $\gamma_1 = R\{Z_1\beta_{11} + (1 - Z_1)\beta_{13}\} + (1 - R)\{Z_2\beta_{12} + (1 - Z_2)\beta_{14}\}$, $\gamma_2 = R\{Z_1\beta_{21} + (1 - Z_1)\beta_{23}\} + (1 - R)\{Z_2\beta_{22} + (1 - Z_2)\beta_{24}\}$. It is not trivial to simulate T^{seq} from (2.6), as the CIFs involved are improper. Here we adopt the simulation strategy used in [Fine and Gray \(1999\)](#), [Cheng et al. \(2009\)](#), and [Beyersmann et al. \(2012\)](#), Sec 5.3. A random variable U is first drawn from Uniform[0,1]. If U is smaller than the asymptote of the CIF, we generate T^{seq} by inverting the CIF. Otherwise, the CIF is not invertible, implying that the cause 2 event occurs first. We assume that the conditional distribution of T^{seq} , given covariates \mathbf{X} and the occurrence of type 2 event, follows $\text{pr}(T^{seq} \leq t \mid \epsilon^{seq} = 2, S = 1, R, Z_1, Z_2, \mathbf{X}) = 1 - \exp\{-t \exp(\gamma_1 X_1 + \gamma_2 X_2)\}$, and simulate T^{seq} from this conditional distribution and let $\epsilon^{seq} = 2$. The true regression coefficients were set as $(\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}) = (0.4, 0.7, 0.5, 0.9)$ and $(\beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}) = (0.3, 0.8, 0.4, 1.2)$.

Assuming that there is no delay between time to the intermediate response and time to the second randomization, the overall survival time is $T = T_1 + ST^{seq}$ with the corresponding cause indicator $\epsilon = (1 - S)\epsilon_1 + S\epsilon^{seq}$. The CIF for cause 1 event at time t for regimen $A_1B_kB'_l$, where $k, l = 1, 2$, can be written as

$$\begin{aligned} & \text{pr}(T_{A_1B_kB'_l} \leq t, \epsilon_{A_1B_kB'_l} = 1) \\ &= p_s \text{pr}(T_1 \leq t, \epsilon_1 = 1 \mid S = 0) + (1 - p_s) \text{pr}(T_1 + T^{seq} \leq t, \epsilon^{seq} = 1 \mid S = 1). \end{aligned} \quad (2.7)$$

Based on these assumptions and the Bayes rule, the first part in (2.7) has the form

$$\text{pr}(T_1 \leq t, \epsilon_1 = 1 \mid S = 0) = \begin{cases} \frac{0.75(1 - e^{-\lambda t})}{1 - (1 - p_r)(1 - p_s)}, & \text{if } t < 0.25, \\ 0.75, & \text{if } t \geq 0.25. \end{cases}$$

For the second part in (2.7),

$$\begin{aligned} \text{pr}(T_1 + T^{seq} \leq t, \epsilon^{seq} = 1 \mid S = 1) &= p_r \text{pr}(T \leq t, \epsilon = 1 \mid A_1 B_k, R = 1, S = 1) \\ &+ (1 - p_r) \text{pr}(T \leq t, \epsilon = 1 \mid A_1 B'_l, R = 0, S = 1). \end{aligned}$$

Thus, we have simulated a cause-1 event time T from subjects following the aforementioned three scenarios with the CIF given in (2.7). Finally, the censoring time, C , was generated from the Exponential distribution with rate 0.1. The observed time $V = \min(T, C)$ and the observed event type indicator $\Delta\epsilon = I(T < C)\epsilon$.

We generated samples with size 400 and repeated 2000 times. For each simulated data, we implemented six models, including the original Fine and Gray model (FG), the fixed weight Fine and Gray model (WFG), the time-varying weight Fine and Gray model (TWFG), the original Scheike Model (SC), the fixed weight Scheike Model (WSC), and the time-varying weight Scheike Model (TWSC). We implemented the WFG and TWFG by solving the score functions as discussed in Sections 2.2.2 and 2.2.3, and then computed the estimated CIF as well as standard deviation based on the influence functions given in Section 2.3. The FG was simply a special case by setting all weights to be 1. For the Scheike models, we assumed time-varying effects for both covariates, and used the cloglog link function to have a proportional hazards model. The implementation of the WSC and TWSC can be completed by solving the score functions for both models, and then using the functional delta method to estimate the variance of CIFs. The naive Scheike model can be simply run by using the R function “comp.risk”. In order to compare all models, the weights for subjects with $S = 0$ were set equal to 1 for the FG and SC.

For each model, we computed the averages of the CIF estimates at different time points with covariates $(X_1, X_2) = (1.5, 1)$. The true CIF values were computed based on (2.7) through numerical integration. We considered two probabilities of response, $p_r = 0.4$ or 0.7 , and presented the results for $A_1 B_1 B'_1$ and $A_1 B_1 B'_2$ regimens in Figure 2.2 over a range of time points. In Tables 2.1 and 2.2, we also listed the mean of estimates (est), the empirical standard deviation ($\tilde{\sigma}$), the mean of estimated standard deviations ($\hat{\sigma}$), and the coverage rate of 95% confidence intervals (Cov) along with the true values at time points 0.225, 0.3, 0.5, 0.75, and 1.

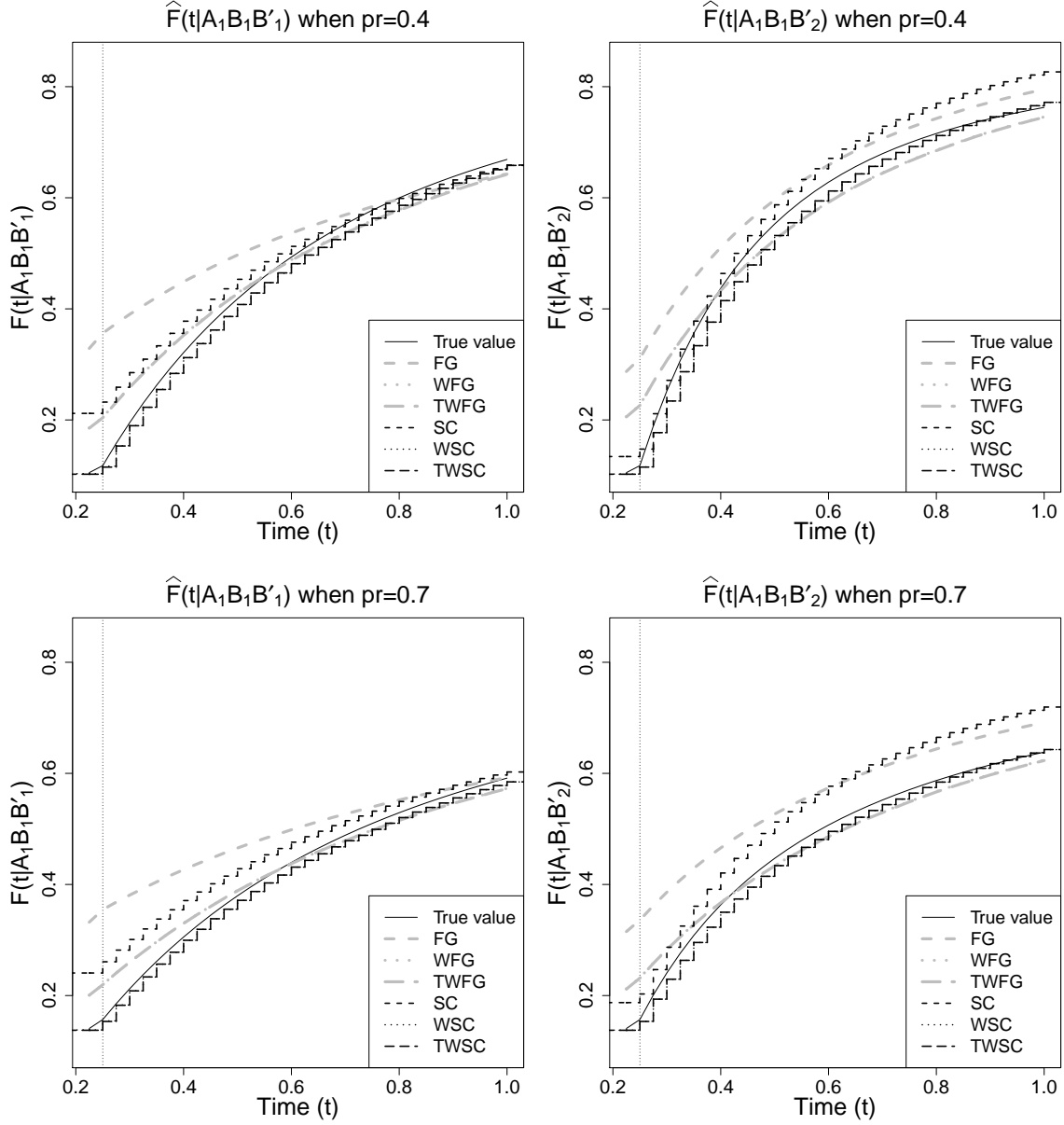


Figure 2.2: The estimated CIFs over time using six models. The black solid line is the true function. Gray lines for Fine-Gray-related models. Black stepwise curves for Scheike-related models. The native methods are dashed lines, the fixed weight methods are dotted lines and the time-varying weight are long dashed lines.

Table 2.1: Simulation results for $A_1B_1B'_1$ and $A_1B_1B'_2$ with $n=400$ and $p_r=0.4$. The time point (time), the method for estimate (method), the true cumulative incidence (true), mean of estimates (est), empirical standard deviation ($\tilde{\sigma}$), mean of estimated standard deviations ($\hat{\sigma}$), coverage rate of 95% confidence intervals (Cov).

p_r	time	method	$A_1B_1B'_1$					$A_1B_1B'_2$				
			true	est	$\tilde{\sigma}$	$\hat{\sigma}$	Cov	true	est	$\tilde{\sigma}$	$\hat{\sigma}$	Cov
0.4	0.225	FG	0.11	0.33	0.06	0.06	0.01	0.11	0.29	0.05	0.05	0.01
		WFG		0.19	0.04	0.04	0.61		0.21	0.05	0.04	0.37
		TWFG		0.19	0.04	0.04	0.57		0.21	0.04	0.04	0.34
		SC		0.21	0.07	0.07	0.78		0.13	0.05	0.05	0.96
		WSC		0.10	0.05	0.04	0.88		0.10	0.05	0.04	0.87
		TWSC		0.10	0.04	0.04	0.88		0.10	0.04	0.04	0.88
	0.3	FG	0.20	0.39	0.07	0.07	0.11	0.25	0.39	0.05	0.05	0.27
		WFG		0.26	0.06	0.06	0.86		0.31	0.06	0.06	0.87
		TWFG		0.26	0.06	0.06	0.86		0.31	0.06	0.06	0.86
		SC		0.29	0.09	0.08	0.86		0.27	0.07	0.07	0.93
		WSC		0.19	0.08	0.07	0.87		0.23	0.08	0.07	0.86
		TWSC		0.19	0.08	0.07	0.87		0.23	0.08	0.07	0.86
	0.5	FG	0.42	0.5	0.07	0.07	0.82	0.55	0.6	0.06	0.06	0.87
		WFG		0.43	0.08	0.08	0.95		0.52	0.07	0.07	0.91
		TWFG		0.43	0.08	0.08	0.94		0.52	0.07	0.07	0.92
		SC		0.45	0.10	0.10	0.93		0.59	0.08	0.08	0.92
		WSC		0.41	0.11	0.11	0.89		0.53	0.09	0.09	0.92
		TWSC		0.41	0.11	0.11	0.90		0.53	0.09	0.09	0.92
	0.75	FG	0.58	0.58	0.08	0.08	0.94	0.7	0.73	0.06	0.06	0.89
		WFG		0.56	0.08	0.08	0.93		0.67	0.07	0.07	0.93
		TWFG		0.56	0.08	0.09	0.93		0.67	0.07	0.07	0.93
		SC		0.58	0.10	0.10	0.93		0.75	0.07	0.07	0.85
		WSC		0.56	0.11	0.11	0.93		0.69	0.09	0.09	0.93
		TWSC		0.56	0.11	0.11	0.93		0.69	0.09	0.09	0.93
	1	FG	0.67	0.64	0.07	0.08	0.94	0.76	0.8	0.05	0.05	0.85
		WFG		0.64	0.08	0.09	0.93		0.75	0.07	0.07	0.94
		TWFG		0.64	0.08	0.09	0.93		0.74	0.07	0.07	0.94
		SC		0.66	0.09	0.10	0.95		0.83	0.07	0.06	0.76
		WSC		0.66	0.10	0.11	0.94		0.77	0.08	0.08	0.91
		TWSC		0.66	0.10	0.11	0.94		0.77	0.08	0.08	0.91

Table 2.2: Simulation results for $A_1B_1B'_1$ and $A_1B_1B'_2$ with $n=400$ and $p_r=0.7$. The time point (time), the method for estimate (method), the true cumulative incidence (true), mean of estimates (est), empirical standard deviation ($\tilde{\sigma}$), mean of estimated standard deviations ($\hat{\sigma}$), coverage rate of 95% confidence intervals (Cov).

p_r	time	method	$A_1B_1B'_1$					$A_1B_1B'_2$				
			true	est	$\tilde{\sigma}$	$\hat{\sigma}$	Cov	true	est	$\tilde{\sigma}$	$\hat{\sigma}$	Cov
0.7	0.225	FG	0.14	0.33	0.06	0.06	0.08	0.14	0.32	0.05	0.05	0.05
		WFG		0.20	0.05	0.05	0.84		0.21	0.05	0.05	0.77
		TWFG		0.20	0.05	0.05	0.83		0.21	0.05	0.05	0.75
		SC		0.24	0.07	0.08	0.83		0.19	0.06	0.06	0.94
		WSC		0.14	0.06	0.06	0.88		0.14	0.06	0.06	0.89
		TWSC		0.14	0.05	0.05	0.89		0.14	0.05	0.05	0.89
	0.3	FG	0.21	0.38	0.07	0.07	0.28	0.24	0.39	0.06	0.06	0.29
		WFG		0.26	0.06	0.06	0.92		0.28	0.06	0.06	0.93
		TWFG		0.26	0.06	0.06	0.92		0.28	0.06	0.06	0.91
		SC		0.30	0.08	0.08	0.88		0.29	0.08	0.08	0.93
		WSC		0.21	0.08	0.08	0.89		0.23	0.08	0.08	0.88
		TWSC		0.21	0.08	0.08	0.89		0.23	0.08	0.08	0.88
	0.5	FG	0.38	0.47	0.07	0.07	0.80	0.45	0.53	0.07	0.07	0.79
		WFG		0.39	0.07	0.08	0.95		0.43	0.07	0.07	0.93
		TWFG		0.39	0.07	0.08	0.94		0.43	0.07	0.07	0.93
		SC		0.43	0.09	0.09	0.93		0.51	0.09	0.09	0.89
		WSC		0.37	0.1	0.1	0.91		0.43	0.1	0.10	0.91
		TWSC		0.37	0.1	0.1	0.92		0.43	0.1	0.10	0.91
	0.75	FG	0.51	0.54	0.08	0.08	0.93	0.57	0.63	0.07	0.07	0.85
		WFG		0.5	0.08	0.08	0.95		0.55	0.08	0.08	0.93
		TWFG		0.5	0.08	0.08	0.94		0.55	0.08	0.08	0.93
		SC		0.53	0.09	0.10	0.94		0.65	0.09	0.09	0.84
		WSC		0.50	0.11	0.11	0.93		0.56	0.10	0.10	0.93
		TWSC		0.50	0.11	0.11	0.93		0.56	0.10	0.10	0.93
	1	FG	0.59	0.59	0.08	0.08	0.95	0.64	0.69	0.07	0.07	0.85
		WFG		0.57	0.08	0.09	0.94		0.62	0.08	0.08	0.94
		TWFG		0.57	0.08	0.09	0.94		0.62	0.08	0.08	0.94
		SC		0.60	0.09	0.10	0.94		0.72	0.08	0.08	0.81
		WSC		0.58	0.10	0.11	0.94		0.64	0.10	0.10	0.93
		TWSC		0.58	0.10	0.11	0.94		0.64	0.10	0.10	0.93

The simulation results show that the naive estimators, FG and SC, tend to overestimate the CIF at early time points but not at later time points. This is as expected because the naive methods assign equal weights to those subjects who failed before the second-randomization (with quick failure) and to those subjects who were further randomized to second-stage treatments. In other words, the naive methods include more subjects who failed earlier. In contrast, the weighted estimators, WFG, WSC, TWFG and TWSC, have much better performance in estimating the CIF, since they all properly up-weigh those subjects going through the second-stage randomization. The Scheike model has better performance than the Fine and Gray model, especially when the CIF is relatively low. However, the WFG and TWFG model still perform reasonably well across time. Moreover, though none of the models are exactly the true models, the WFG, TWFG, WSC and TWSC still provide reliable estimation of the CIF. Consequently, the fixed and time-varying weight estimators are more reliable methods than the naive ones in finding an optimal DTR from a two-stage randomized trial.

2.5 ANALYSIS OF CHILDREN’S NEUROBLASTOMA STUDY

We now revisit our motivating example of the neuroblastoma study that was conducted by the Children’s Cancer Group between 1991 and 1996. Neuroblastoma is a type of cancer that starts in early nerve cells of the sympathetic nervous system and occurs most often in infancy and young children. Children with high-risk neuroblastoma have high recurrence and poor survival rates (www.cancer.org/cancer/neuroblastoma). Thus, an important clinical question is how to stop or delay disease progression and thus improve survival, by providing an optimal regimen to patients according to their states. In this section, we apply our methods to evaluate the preventive effect on disease progression of myeloablative chemotherapy and radiotherapy plus purged autologous bone marrow transplantation (ABMT) over intensive chemotherapy (Chemo) alone, followed by subsequent treatment with 13-*cis*-retinoic acid (*cis*-RA) or no further treatment (no RA) for children with high-risk neuroblastoma ([Matthay et al., 2009](#)). The study adopted a two-stage SMART design. After receiving

an induction chemotherapy, 379 eligible children without progressive disease participated in the first-stage randomization, where 189 children were assigned to ABMT and 190 children were assigned to Chemo. Those children, who did not develop progressive disease after the initial treatment and were willing to be further randomized, were defined as Responders and, subsequently randomized to receive either *cis*-RA or no RA. At the second stage, 50 of 98 ABMT responders and 52 of 105 Chemo responders received *cis*-RA. For simplicity, we referred to those children who did not have the second stage randomization as non-responders. Thus, four possible regimens could be constructed for this study: (i) treating with ABMT followed by *cis*-RA if subjects responded and no further therapy if subjects did not respond (ABMT/*cis*-RA); (ii) treating with ABMT followed by no RA if subjects responded and no further therapy if subjects did not respond (ABMT/no RA); and (iii) Chemo/*cis*-RA and (iv) Chemo/no RA were defined similarly.

During the study a total of 269 children developed progressive disease, with 134 occurring in non-responders, a total of 23 children died before they developed the disease, with 22 in non-responders, and a total of 87 children were right censored, with 20 in non-responders. Therefore, the event of interest, the time to disease progression, could not be observed after death, which is a competing event, and the CIF is used to describe cumulative risks of disease progression in the presence of death. Furthermore, an interesting feature of the data is that the response was defined as no disease progression, and the time to response was closely related to our event of interest which is disease progression. As a result, the time-varying weight methods are not applicable; see [Yavuz et al. \(2018\)](#) for more details. Therefore, only the fixed weight methods WFG and WSC can be applied to this dataset.

Following [Matthay et al. \(2009\)](#), we considered five potential risk factors, age (Age), disease stage (Stage4dx), ferritin (Ferritindx), MYCN status (MYCNdx) and bone metastases (Bonesdx). Tumor pathology was not considered due to a very unbalanced sample size (6 vs.120 in ABMT and 9 vs.128 in Chemo). In the analysis, we treated Age as a continuous variable, and included the rest of covariates as dichotomous variables using [Matthay et al. \(2009\)](#)'s definition. Following [Matthay et al. \(2009\)](#), we excluded the missing values and used the complete data with a total of 260 children, with 120 in ABMT and 140 in Chemo. In the complete data, a total of 177 children developed progressive disease, with 95 occurring

in non-responders, a total of 14 children died before they developed the disease, with 14 in non-responders, and a total of 69 children were right censored, with 17 in non-responders.

In order to compare with [Matthay et al. \(2009\)](#), we illustrated our methods by focusing on the AMBT/cis-RA regimen, and applying the WSC model to examine if any time-varying effect exists. The estimated time-varying coefficients are given in Figure 2.3. The formal Kolmogorov-Smirnov test and the Cramer-von Mises test for time-varying coefficients are summarized in Table 2.3, which suggest that none of the covariates have a time-varying effect. Therefore, presenting the fixed-weight Fine and Gray model as the final model for the AMBT/cis-RA regimen is reliable.

Moreover, to compare with other regimens, we examined the time-varying effects of the five covariates using the Schemke model for the rest of three regimens. Only Age is significant with $p\text{-value}=0.02$ in the Chemo/No RA regimen by the Cramer-von Mises test. However, Age is not significant with $p\text{-value}=0.07$ by using the Kolmogorov-Smirnov test. Considering multiple comparisons involved in testing five covariates for four regimens, it is reasonable to assume constant covariate effects, and thus to apply the fixed-weight Fine and Gray model to each of the four DTRs. The p -values for testing the significance of covariate effects in the final models are given in Table 2.4. The results show that Ferritindx, MYCNdx and Age are significant in three of the four regimens, Stage4dx is significant in only one of the four regimens, and Bonedx is not significant for all regimens. The estimated coefficients suggest that higher levels of ferritin and MYCN amplification are associated with faster disease progression. Despite that the outcomes in [Matthay et al. \(2009\)](#) are not the same as our outcome of interest, and the subgroups included in the two analyses are different, our method has identified the same set of important covariates as those listed in [Matthay et al. \(2009\)](#).

To compare the CIFs of progressive disease over time of the four regimens with various covariate effects, we present the CIF estimates obtained by fitting the WFG model for the four regimens in Figure 2.4, for Ferritindx = 0 or 1, and MYCNdx = 0 or 1, while setting Age = 3 (the median age in the data), Stage4dx = 0 and Bonedx = 0. From Figure 2.4, patients with higher level of Ferritindx or MYCN gene copy were more likely to experience progressive disease across the four regimens, which is consistent with the estimated coefficients in Table

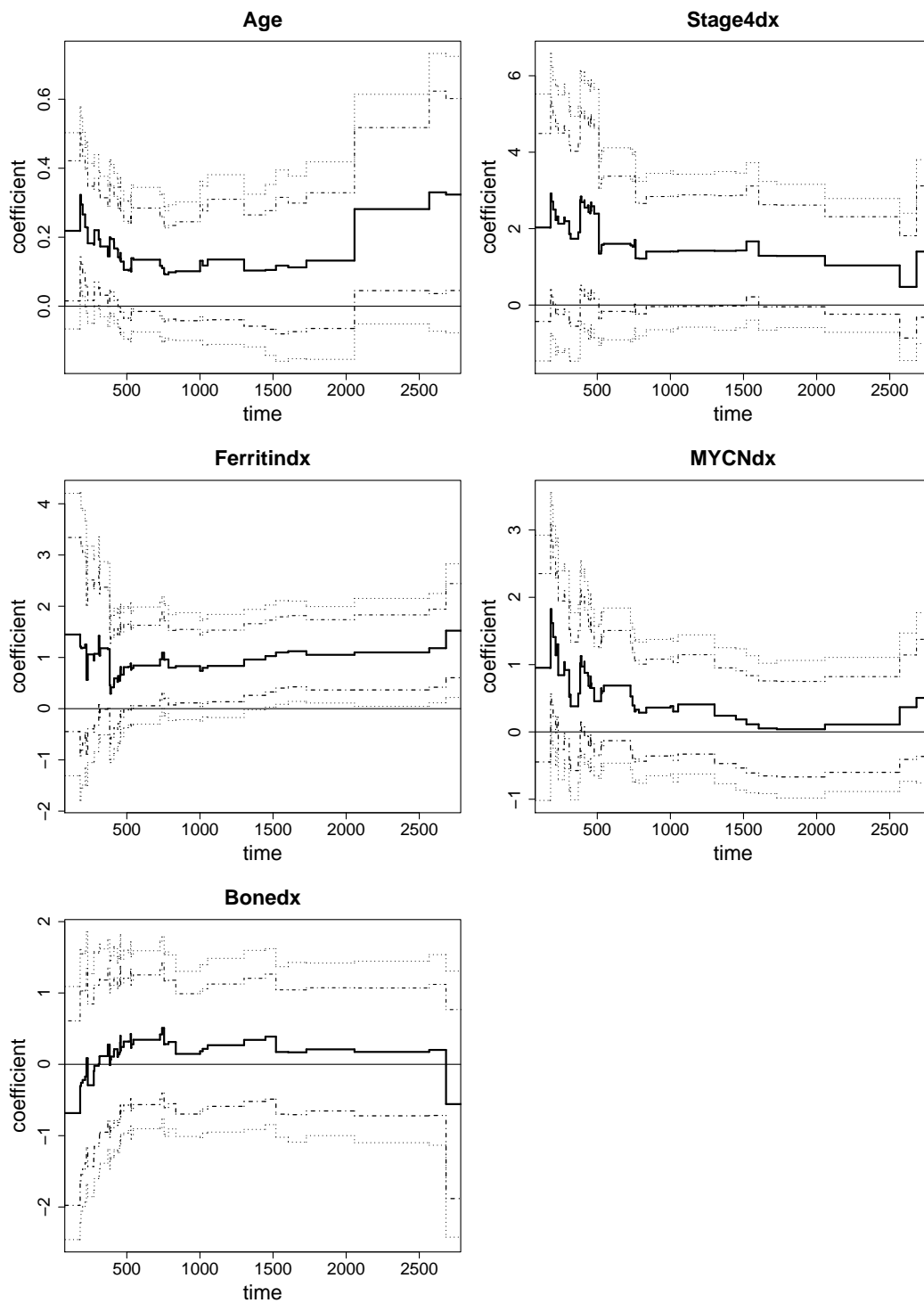


Figure 2.3: The estimated coefficient for regimen ABMT/cis-RA using the fixed weight Scheike model. The solid lines are estimates along with their confidence intervals (dashed lines) and confident bands (dotted lines).

Table 2.3: P-values for testing “time-varying effect” for each variable, using the Kolmogorov-Smirnov test and the Cramer-von Mises test.

Test	P-value of testing time-varying coefficient				
	Age	Stage4dx	Ferritmdx	MYCNdx	Bonedx
Kolmogorov-Smirnov	0.661	0.307	0.566	0.090	0.735
Cramer-von Mises	0.174	0.777	0.276	0.334	0.287

Table 2.4: The estimated coefficients from the WFG model for the four regimens. The estimate $\hat{\beta}$, the estimate of standard deviation ($\hat{\sigma}(\hat{\beta})$) and the P-value for testing $\beta = 0$.

Regimen	Estimates/p-value	Covariate				
		Age	Stage4dx	Ferritindx	MYCNdx	Bonedx
ABMT/cis-RA	$\hat{\beta}$ ($\hat{\sigma}(\hat{\beta})$)	0.17(0.06)	0.97(0.45)	0.97(0.33)	0.51(0.32)	0.12(0.35)
	p-value	0.005	0.031	0.003	0.110	0.737
ABMT/no RA	$\hat{\beta}$ ($\hat{\sigma}(\hat{\beta})$)	0.19(0.05)	0.46(0.45)	0.07(0.31)	0.79(0.38)	0.22(0.35)
	p-value	<0.001	0.312	0.828	0.039	0.532
Chemo/cis-RA	$\hat{\beta}$ ($\hat{\sigma}(\hat{\beta})$)	0.12(0.05)	0.67(0.42)	0.55(0.26)	0.73(0.27)	0.51(0.26)
	p-value	0.034	0.113	0.037	0.007	0.053
Chemo/no RA	$\hat{\beta}$ ($\hat{\sigma}(\hat{\beta})$)	0.06(0.05)	0.37(0.44)	0.68(0.26)	1.11(0.27)	0.02(0.24)
	p-value	0.287	0.407	0.009	< 0.001	0.942

2.4. Figure 2.4 also suggests that the Chemo/no RA regimen seems to be the worst regimen for children with high risk Neuroblastoma, whereas the ABMT/*cis*-RA regimen may be the optimal regimen among the four. For children with high ferritin level and no MYCN gene copy, ABMT/*cis*-RA and ABMT/no RA regimens seem to be comparable with negligible differences, and they both appear to perform better than the other two.

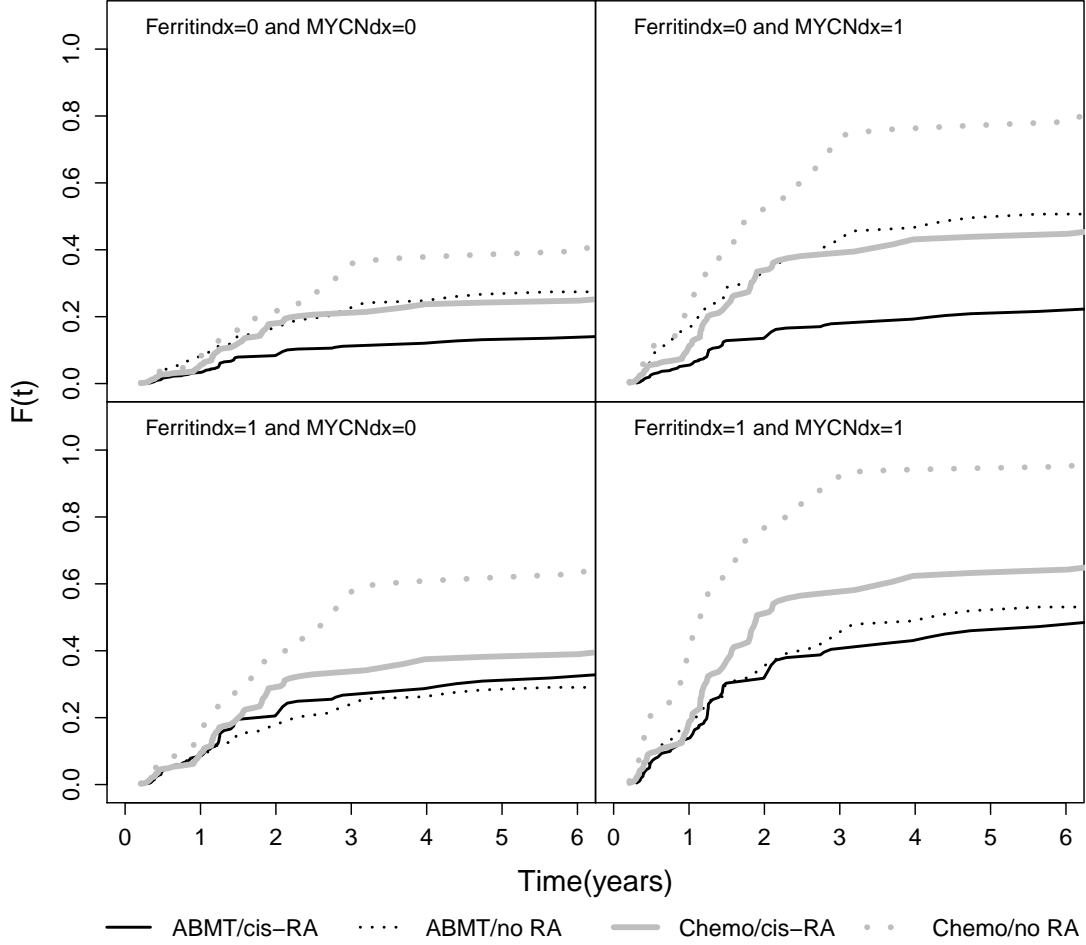


Figure 2.4: The estimated CIFs for the four regimens obtained by using the WFG method with four cases while controlling for Age=3 years, Stage4dx=0 and Bonedx=0. The plots in the upper row are for Ferritindx=0 and the plots in the lower row are for Ferritindx=1. The plots in the left column are for MYCNdx=0 and the plots in the right column are for MYCNdx=1.

2.6 DISCUSSION

Patient heterogeneity is of great clinical importance from a clinical trial perspective. If subjects were to follow a specific DTR, they may wonder how well they would fare from this specific treatment strategy given their own clinical characteristics. In this chapter, we have focused on the direct modeling of covariate effects on a specific DTR. We have demonstrated that the inverse-probability-weighting method can be used to extend some commonly used regression models for competing-risk data to a two-stage randomization setting. The Fine and Gray and Scheike models were used as examples, though our methods can be readily applied to other models, such as the multinomial logistic model ([Gerds et al., 2012](#)). Our simulations show that the resulting weighted estimators of the CIF are still reasonably accurate, even though the underlying Fine and Gray or Scheike model may be misspecified. Therefore, we provide convenient and reliable methods to evaluate covariate effects on the CIF. The proper modeling of covariate effects on various DTRs will facilitate selection of the optimal treatment strategy for a subject with specific characteristics.

3.0 QUANTILE ASOCIATION MODEL FOR BIVARIATE SURVIVAL DATA

3.1 INTRODUCTION

The association between two failure times is often of interest in familial studies, finance and biomedical research. For example, in an atherosclerosis study, two diseases, myocardial infarction and stroke, are probably associated with each other. Understanding their association may help prevent the occurrence of one event, once the other event is observed. Another example is age-related macular degeneration (AMD), which is a leading cause of vision loss in developed countries ([Swaroop et al., 2009](#)). A patient who was identified to have AMD in one eye may have a higher risk of developing AMD in the other eye. Several global dependence measures have been developed to measure the strength of association between the dependent pairs. [Oakes \(1982, 2008\)](#) proposed a nonparameteric estimator of Kendall's tau under the presence of the censoring. [Wang and Wells \(2000\)](#) further introduced other estimators of Kendall's tau using V-statistics. [Lakhal et al. \(2009\)](#) adopted the inverse probability censoring weighted method in the estimation of Kendall's tau. In addition, [Hsu and Prentice \(1996\)](#) proposed an estimator for the correlation between two cumulative variates using a nonparameteric method.

Global association measures are appealing for their ease of interpretation. However, they cannot capture the local association structure which may vary over time. There are extensive works on local associations. One approach to quantifying local association is to analyze the bivariate survival data by a frailty model, or more generally, under the copula framework, where copula models allow time-dependent association between two failure times. [Oakes \(1989\)](#) showed that the Clayton model ([Clayton, 1978](#)) can be cast under

the frailty framework. [Anderson et al. \(1992\)](#) considered the time-dependent conditional expected residual life and conditional probability to quantify time-dependent association in bivariate survival data under the proportional hazard frailty model. [Shih and Louis \(1995\)](#) proposed two two-stage estimation procedures for the association parameters in copula models. Other time-varying measures include a martingale covariance function for two failure times ([Prentice and Cai, 1992](#)), a piecewise constant cross hazard ratio ([Nan et al., 2006](#)), and a time-dependent cross ratio ([Hu and Nan, 2011](#)), among others.

In the analysis of association, it is of interest to investigate how risk factors are related to the local association structure between two event times. Conditional association tends to be more reasonable because it can evaluate the important factors, as well as eliminate potential confounders. In the AMD example, age, family history and smoking status are considered as possible risk factors for the development of AMD. They may also influence how the onset times of the AMD are related with one another from the same subject. By identifying those patients with stronger local association, researchers may provide effective treatments for them once they have developed AMD in one eye to prevent the development of AMD in the other eye. In decades, much research focuses on adjusting for covariate effects on marginal distributions, but not directly on the association. For instance, [Zeng et al. \(2009\)](#) formed a general transformation for the cumulative hazard function in a gamma frailty model, while considering covariates for the marginal cumulative hazard functions. [Li et al. \(2016\)](#) proposed an association model based on the odds ratio for quantiles, and considered the covariate effects on the marginal distributions only. However, few works target conditional association, as it is challenging to evaluate covariate effects on the strength of local association. [Yan and Fine \(2005\)](#) proposed a functional association regression model on a temporal process with time-varying coefficient effects, though the temporal association may be affected by the assumed marginal distributions.

In this work, we propose a conditional association model for bivariate survival data, by adopting a novel quantile-specific association measure – quantile odds ratio (qor) as proposed in [Li et al. \(2014\)](#). The qor is independent of the marginal distributions, invariant to monotone transformations, and insensitive to outliers. [Li et al. \(2014\)](#) utilized existing quantile regression models to allow covariate effects on marginal quantiles, and developed

regression models for the *gor* for completely observed bivariate outcomes. For bivariate survival data, [Li et al. \(2016\)](#) successfully explored the quantile association through the *gor* in the copula framework, and proposed two estimators of the quantile association by using nonparametric and semi-parametric methods, respectively. Although [Li et al. \(2016\)](#) considered the covariate effects in the estimation of the quantile association, they assumed that covariate effects influence the quantile association via marginal quantiles only, which may not be true in bivariate survival data. Therefore, we propose a quantile association model for censored pairs and allow covariate effects on both marginal distributions and the local association structure. More specifically, we adopt the censored quantile regression models for marginal quantiles, and construct a quantile-based regression model for the transformed *gor* in bivariate survival data.

Quantile regression ([Koenker and Bassett, 1978](#)) is attractive in studying dynamic effects of covariates on an outcome, because it allows researchers to assess covariate effects across different quantiles of the outcome, and regression coefficients are easy to interpret. Quantile regression has been well extended to univariate survival data under different scenarios, such as survival data with independent censoring ([Portnoy, 2003](#); [Peng and Huang, 2008](#); [Koenker et al., 2008](#)), competing risks data ([Peng and Fine, 2009](#)), left-truncated semi-competing risks data ([Li and Peng, 2011](#)), among others. Meanwhile in the past decade, several censored quantile models were developed for correlated survival data. For example, [Yin and Cai \(2005\)](#) proposed a quantile regression model for multivariate failure times with an independent working covariance matrix, and estimated parameters by generalized estimating equations. [Ji et al. \(2014\)](#) developed quantile models for marginal failure times and handled dependent censoring times through a copula model for the joint distribution. [Li and Peng \(2015\)](#) further proposed quantile regression approaches to deal with dependent censoring and semi-competing risks censoring. Compared to the abundant extensions of quantile regression for univariate survival data, there is very limited work studying covariate effects on quantile association for bivariate survival data.

Therefore, we propose a conditional quantile association model that allows covariate effects on both the marginal distributions and the association structure. Under the random censoring assumption, we first adopt the censored quantile regressions for marginal condi-

tional quantiles. We then propose a model to estimate the effects of the covariates on the conditional qor , through the relationship between qor and the conditional copula function. The estimation of covariance matrices is often tricky for quantile regression and quantile association analyses due to the unsmooth objective functions in the optimization process. We thus extend an idea of the induced smoothing procedure (Brown and Wang, 2005) to estimate the influence functions for our proposed estimators, and propose an algorithm to obtain a consistent estimator for the covariance matrix of the proposed estimators. Our proposed method explicitly addresses the presence of right censoring and greatly expands the application of the method in Li et al. (2014) to time-to-event types of data. The rest of this chapter is organized as follows. We propose our conditional quantile association model and estimating equations in Sections 3.2.1 and 3.2.2. The asymptotic properties for the coefficient estimates and the covariance estimation are given in Sections 3.2.3 and 3.2.4. We present numerical simulations for the proposed method and procedure in Section 3.3, and apply to an AMD dataset in Section 3.4. Finally, some discussions are given in Section 3.5.

3.2 METHOD

3.2.1 Bivariate Survival Data and Models

To begin, we introduce necessary notation for bivariate survival data with covariates. Let (T_1, T_2) be a vector of bivariate survival times, and (C_1, C_2) be the corresponding vector of bivariate right censoring times. Define $Y_j = \min(T_j, C_j)$, $\delta_j = I(T_j \leq C_j)$, $j = 1, 2$. Let \mathbf{Z}_j denote the covariate vector, which includes 1 as the first element and time-independent covariates, that is relevant to T_j , $j = 1, 2$, and let \mathbf{Z}_3 denote the covariate vector that is directly related to the association between T_1 and T_2 . Define \mathbf{Z} as a vector that consists all p covariates in \mathbf{Z}_1 , \mathbf{Z}_2 , and \mathbf{Z}_3 . In the presence of independent censoring, the observed bivariate survival data consist of n i.i.d. replicates of $\{Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i}, \mathbf{Z}_i\}_{i=1}^n$.

For $j = 1, 2$, define $F_j(t|\mathbf{Z}_j) = \Pr(T_j \leq t|\mathbf{Z}_j)$ as the marginal conditional cumulative function of T_j , and $Q_j(u|\mathbf{Z}_j) = \inf\{t : F_j(t|\mathbf{Z}_j) \geq u\}$, $u \in (0, 1)$ as the corresponding

marginal conditional quantile function. Let $H(t_1, t_2|\mathbf{Z}) = \Pr(T_1 \leq t_1, T_2 \leq t_2|\mathbf{Z})$ be the conditional bivariate cumulative distribution function of (T_1, T_2) . The conditional copula function is defined as

$$\mathcal{C}(\boldsymbol{\tau}|\mathbf{Z}) := \Pr(T_1 \leq Q_1(\tau_1|\mathbf{Z}_1), T_2 \leq Q_2(\tau_2|\mathbf{Z}_2)|\mathbf{Z}) = H\{Q_1(\tau_1|\mathbf{Z}_1), Q_2(\tau_2|\mathbf{Z}_2)|\mathbf{Z}\},$$

where $\boldsymbol{\tau} = (\tau_1, \tau_2) \in (0, 1)^2$. To simplify the notation, we simply use $F_j(T_j|\mathbf{Z})$ to denote $F_j(T_j|\mathbf{Z}_j)$, for $j = 1, 2$, with the understanding that not all covariates in \mathbf{Z} are significantly related to T_j . Thus, $H\{Q_1(\tau_1|\mathbf{Z}), Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\} \stackrel{d}{=} H\{Q_1(\tau_1|\mathbf{Z}_1), Q_2(\tau_2|\mathbf{Z}_2)|\mathbf{Z}\}$.

In this study, we adopt a novel quantile association measure, quantile-specific odds ratio (*qor*), that was proposed by Li et al. (2014), where

$$\begin{aligned} qor(\boldsymbol{\tau}|\mathbf{Z}) &= \frac{\text{odds}\{T_1 \leq Q_1(\tau_1|\mathbf{Z})|T_2 \leq Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\}}{\text{odds}\{T_1 \leq Q_1(\tau_1|\mathbf{Z})|T_2 > Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\}} \\ &= \frac{\text{odds}\{T_1 > Q_1(\tau_1|\mathbf{Z})|T_2 > Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\}}{\text{odds}\{T_1 > Q_1(\tau_1|\mathbf{Z})|T_2 \leq Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\}} \\ &= \frac{\Pr\{T_1 \leq Q_1(\tau_1|\mathbf{Z}), T_2 \leq Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\} \times \Pr\{T_1 > Q_1(\tau_1|\mathbf{Z}), T_2 > Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\}}{\Pr\{T_1 \leq Q_1(\tau_1|\mathbf{Z}), T_2 > Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\} \times \Pr\{T_1 > Q_1(\tau_1|\mathbf{Z}), T_2 \leq Q_2(\tau_2|\mathbf{Z})|\mathbf{Z}\}}. \end{aligned} \quad (3.1)$$

The *qor* represents the odds that the first event occurs before (after) its quantile $Q_1(\tau_1)$ given that the second event occurred before (after) its quantile $Q_2(\tau_2)$, compared to the odds that the first event occurs before (after) its quantile $Q_1(\tau_1)$ given that the second event occurred after (before) its quantile $Q_2(\tau_2)$ (Li et al., 2014, 2016). Expressed as an odds ratio, it is easy to interpret the relationship between two event times based on the *qor*. If there exists a positive (negative) association between T_1 and T_2 , the *qor* is greater (less) than 1. If they are independent, then the *qor* is equal to 1. Under different copula models, the *qor* changes with τ , except for the Plackett copula under which the *qor* stays constant; see Li et al. (2014) for more details.

To connect the conditional copula function with the *qor*, we can express $\mathcal{C}(\boldsymbol{\tau}|\mathbf{Z})$ as a function of the *qor*, $\chi(qor(\boldsymbol{\tau}|\mathbf{Z}), \boldsymbol{\tau})$, where

$$\chi(y; \boldsymbol{\tau}) := \begin{cases} \frac{\tau_1 + \tau_2}{2} + \frac{1 - \sqrt{(y-1)^2(\tau_1 - \tau_2)^2 + 2(y-1)(\tau_1 + \tau_2 - 2\tau_1\tau_2) + 1}}{2(y-1)} & \text{if } y \neq 1; \\ \tau_1\tau_2 & \text{if } y = 1, \end{cases} \quad (3.2)$$

and $\lim_{y \rightarrow 0+} \chi(y; \boldsymbol{\tau}) = \max(0, \tau_1 + \tau_2 - 1)$ and $\lim_{y \rightarrow \infty} \chi(y; \boldsymbol{\tau}) = \min(\tau_1, \tau_2)$. It is simple to show that the conditional copula function has a monotone relationship with the *gor*.

In this study, we model the effects of \mathbf{Z} on the marginal distributions, by adopting the censored quantile regression model (Portnoy, 2003; Peng and Huang, 2008; Koenker et al., 2008), which assumes that

$$Q_j(\tau_j | \mathbf{Z}) = g_j\{\mathbf{Z}^T \boldsymbol{\beta}_{j0}(\tau_j)\}, \tau_j \in (0, \tau_{U_j}), \quad (3.3)$$

where $\boldsymbol{\beta}_{j0}(\tau_j)$ is a $p \times 1$ vector of unknown coefficients, τ_{U_j} is the maximum quantile level that is estimable from the censored data, and $g_j(\cdot)$ is a known monotone link function, for $j = 1, 2$.

To model the local conditional association, we assume that

$$\log \text{gor}(\boldsymbol{\tau} | \mathbf{Z}) = \mathbf{Z}^T \boldsymbol{\gamma}_0(\boldsymbol{\tau}), \quad (3.4)$$

where $\boldsymbol{\gamma}_0(\boldsymbol{\tau}) = \{\gamma_0^{(0)}(\boldsymbol{\tau}), \gamma_0^{(1)}(\boldsymbol{\tau}), \dots, \gamma_0^{(p-1)}(\boldsymbol{\tau})\}$ is a $p \times 1$ vector of coefficients. $\gamma_0^{(0)}(\boldsymbol{\tau})$ corresponds to the baseline $\log \text{gor}(\cdot)$ when all covariates are set to zero. The absolute value of $\gamma_0^{(k)}(\boldsymbol{\tau})$ and the sign of $\gamma_0^{(k)}(\boldsymbol{\tau})$ represent the magnitude and the direction of the changes in the local association at the τ -th quantiles, when the k^{th} covariate increases, with $k = 1, \dots, p-1$. Under this structure, the conditional copula function has a form, $\mathcal{C}(\boldsymbol{\tau} | \mathbf{Z}) = \chi[\exp\{\mathbf{Z}^T \boldsymbol{\gamma}_0(\boldsymbol{\tau})\}; \boldsymbol{\tau}]$. Again in this study, we simply use the same \mathbf{Z} for both marginal models and the local association model for the brevity. In fact, different sets of covariates are allowed in the models (3.3) and (3.4).

3.2.2 Estimating Equations

Before evaluating the association coefficients, $\boldsymbol{\gamma}_0(\boldsymbol{\tau})$, we need to first estimate the unknown parameters, $\boldsymbol{\beta}_{j0}(\tau_j)$, in the marginal censored quantile models. Without loss of generality, we adopt Peng and Fine (2009)'s methods which uses the inverse probability of censoring

weighting (IPCW) technique to modify the standard quantile regression model in the estimation equation. More specifically, let $G_j(t|\mathbf{Z})$ be the survival function of C_j given \mathbf{Z} , $j = 1, 2$. The estimating equation for the true parameters, $\beta_{j0}(\tau_j)$, is,

$$S_{nj}(\mathbf{b}_j; \tau_j) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i \left[\frac{I\{Y_{ji} \leq g_j(\mathbf{Z}^T \mathbf{b}_j)\} \delta_{ji}}{\hat{G}_j(Y_{ji})} - \tau_j \right],$$

where $\hat{G}_j(\cdot)$ denotes a consistent estimator for G_j , such as the Kaplan-Meier estimator, $j = 1, 2$. In practice, we focus on a pre-specified region of $\boldsymbol{\tau} \in \mathbb{D}$, where \mathbb{D} is a subset of $(0, \tau_{U_1}] \times (0, \tau_{U_2}]$. Under mild regularity conditions, it has been shown that $S_{nj}(\mathbf{b}_j; \tau_j)$ is asymptotically mean zero at the true parameters β_{j0} , and solving the estimating equation $S_{nj}(\mathbf{b}_j; \tau_j) = 0$ can be transformed into optimizing a L_1 -type convex function. Therefore, despite that $S_{nj}(\mathbf{b}_j; \tau_j)$ is not smooth, the solution to $S_{nj}(\mathbf{b}_j; \tau_j) = 0$ still can be obtained by minimizing the L_1 -type convex function (Peng and Fine, 2009). We use the existing software package, such as $rq()$ function in R package *quantreg*, to obtain the estimators $\hat{\beta}_j(\tau_j)$ and the corresponding quantile estimators $\hat{Q}_j(\tau_j|\mathbf{Z}) = g_j\{\mathbf{Z}^T \hat{\beta}_j(\tau_j)\}$ for $j = 1, 2$.

We now consider the main objective of this study of evaluating the quantile association effects, $\gamma_0(\boldsymbol{\tau})$, based on bivariate survival data. For complete data,

$$E\{I(T_1 \leq Q_1(\tau_1|\mathbf{Z}), T_2 \leq Q_2(\tau_2|\mathbf{Z})|\mathbf{Z})\} = \mathcal{C}(\boldsymbol{\tau}|\mathbf{Z}).$$

For bivariate survival data, we adapt a commonly used technique based on the IPCW to account for censored observations. Under the assumption that the censoring C_j is conditionally independent of T_j given \mathbf{Z} , we have

$$\begin{aligned} & E\left\{ \frac{I(Y_1 \leq t_1, Y_2 \leq t_2) \delta_1 \delta_2}{G(Y_1, Y_2)} \middle| \mathbf{Z} \right\} \\ &= E\left[E\left\{ \frac{I(T_1 \leq t_1, T_2 \leq t_2) I(T_1 \leq C_1) I(T_2 \leq C_2)}{G(Y_1, Y_2)} \middle| T_1, T_2, \mathbf{Z} \right\} \middle| \mathbf{Z} \right] \\ &= E\left\{ \frac{I(T_1 \leq t_1, T_2 \leq t_2) G(T_1, T_2)}{G(T_1, T_2)} \middle| \mathbf{Z} \right\} \\ &= \Pr(T_1 \leq t_1, T_2 \leq t_2 | \mathbf{Z}) = H(t_1, t_2 | \mathbf{Z}), \end{aligned}$$

where $G(t_1, t_2) = \Pr(C_1 > t_1, C_2 > t_2)$. Let $\hat{G}(t_1, t_2)$ be a consistent estimator of $G(t_1, t_2)$.

We can show that

$$\begin{aligned} E \left[\frac{I\{Y_1 \leq Q_1(\tau_1|\mathbf{Z}), Y_2 \leq Q_2(\tau_2|\mathbf{Z})\} \delta_1 \delta_2}{\hat{G}(Y_1, Y_2)} \middle| \mathbf{Z} \right] &= \Pr(T_1 \leq Q_1(\tau_1|\mathbf{Z}), T_2 \leq Q_2(\tau_2|\mathbf{Z}) | \mathbf{Z}) + o(1) \\ &= \mathcal{C}(\boldsymbol{\tau}|\mathbf{Z}) + o(1) = \chi\{qor(\boldsymbol{\tau}|\mathbf{Z}), \boldsymbol{\tau}\} + o(1). \end{aligned}$$

Along with the consistent estimators of $Q_j(\tau_j|\mathbf{Z})$, $\hat{Q}_j(\tau_j|\mathbf{Z})$, from the marginal quantile regression, and under the assumed conditional association effects model (3.4), we propose the following estimating equation to estimate $\boldsymbol{\gamma}_0(\boldsymbol{\tau})$:

$$\mathbf{W}_n^{\hat{G}}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \boldsymbol{\gamma}; \boldsymbol{\tau}) = \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \left[\frac{I\{Y_{1i} \leq \hat{Q}_1(\tau_1|\mathbf{Z}), Y_{2i} \leq \hat{Q}_2(\tau_2|\mathbf{Z})\} \delta_{1i} \delta_{2i}}{\hat{G}(Y_{1i}, Y_{2i})} - \chi\{\exp(\mathbf{Z}_i^T \boldsymbol{\gamma}); \boldsymbol{\tau}\} \right] = 0,$$

where $\hat{Q}_j(\tau_j|\mathbf{Z}) = g_j\{\mathbf{Z}^T \hat{\boldsymbol{\beta}}_j(\tau_j)\}$ for $j = 1, 2$. For a fixed $\boldsymbol{\tau}$, $\mathbf{W}_n^{\hat{G}}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \boldsymbol{\gamma}; \boldsymbol{\tau})$ is smooth in $\boldsymbol{\gamma}$. Let $\chi'(\cdot; \boldsymbol{\tau})$ be the derivative of $\chi(\cdot; \boldsymbol{\tau})$. $\chi'(y, \boldsymbol{\tau})$ can be shown to be positive for $y \in \mathcal{R}$. Then, $\partial \mathbf{W}_n^{\hat{G}}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \boldsymbol{\gamma}; \boldsymbol{\tau}) / \partial \boldsymbol{\gamma} = -n^{-1} \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T \exp(\mathbf{Z}_i^T \boldsymbol{\gamma}) \chi'\{\exp(\mathbf{Z}_i^T \boldsymbol{\gamma}); \boldsymbol{\tau}\}$ exists, and is a negative definite matrix. This ensures a unique solution to $\mathbf{W}_n^{\hat{G}}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \boldsymbol{\gamma}; \boldsymbol{\tau}) = 0$, which can be found by using the Newton-Raphson algorithm that is implemented by the *multiroot()* function in the R package *rootSolve*. There are a variety of methods to estimating $G(y_1, y_2)$. In the AMD study, we assume the univariate censoring mechanism for both eyes, where $G(y_1, y_2) = \Pr\{C > \max(y_1, y_2)\}$, and adopt the Kaplan-Meier estimator on the basis of $\{\max(Y_{1i}, Y_{2i}), 1 - \delta_{1i} \delta_{2i}\}_{i=1}^n$. For the more general bivariate censoring, consistent estimators such as the [Prentice and Cai \(1992\)](#) method can be used to estimate $G(y_1, y_2)$ on the basis of $\{Y_{1i}, 1 - \delta_{1i}, Y_{2i}, 1 - \delta_{2i}\}_{i=1}^n$.

3.2.3 Asymptotic Properties

In this subsection, we establish the uniform consistency and weak convergence of the proposed estimator $\hat{\gamma}_0(\boldsymbol{\tau})$ for $\boldsymbol{\tau} \in \mathbb{D}$. We first state some notation and the regularity conditions. For a vector \mathbf{u} , define $\mathbf{u}^{\otimes 2} = \mathbf{u}\mathbf{u}^T$ and $\|\mathbf{u}\|$ as its Euclidean norm. We use $\text{eigmin}(\mathbf{A})$ to denote the minimal eigenvalue of a square matrix \mathbf{A} . Let $f_j(t|\mathbf{z}) = dF_j(t|\mathbf{z})/dt$, and $h_j(t_1, t_2|\mathbf{Z}) = \partial H(t_1, t_2|\mathbf{Z})/\partial t_j$, for $j = 1, 2$. Let $\mathbf{A}_j(\mathbf{b}_j) = E[\mathbf{Z}^{\otimes 2} f_j\{g_j(\mathbf{Z}^T \mathbf{b}_j)|\mathbf{Z}\}]$ and $\mathbf{P}_j(\mathbf{b}_1, \mathbf{b}_2) = E[\mathbf{Z}^{\otimes 2} h_j\{g_1(\mathbf{Z}^T \mathbf{b}_1), g_2(\mathbf{Z}^T \mathbf{b}_2)|\mathbf{Z}\} g'_j(\mathbf{Z}^T \mathbf{b}_j)]$, where $g'_j(u) = dg_j(u)/du$. Denote $\mathbf{J}(\boldsymbol{\gamma}; \boldsymbol{\tau}) = E[\mathbf{Z}^{\otimes 2} \chi' \{\exp(\mathbf{Z}^T \boldsymbol{\gamma})\} \exp(\mathbf{Z}_i^T \boldsymbol{\gamma})]$, where $\chi'(u) = d\chi(u)/du$. The required regularity conditions are listed below:

- C1. \mathbf{Z} is uniformly bounded, that is, $\sup_i \|\mathbf{Z}_i\| < \infty$ for $i = 1, \dots, n$.
- C2. There exists $k_j > 0$ such that $\Pr(C_j = k_j) > 0$ and $\Pr(C_j > k_j) = 0$, for $j = 1, 2$.
Moreover, there exists $\delta > 0$ such that $\Pr(C_1 \geq c_1, C_2 \geq c_2) > \delta > 0$ for any $c_j \leq k_j, j = 1, 2$.
- C3. (i) $f_j(t|\mathbf{z})$ is bounded uniformly in t and \mathbf{z} , for $j = 1, 2$; (ii) $\beta_{j0}(\tau_j)$ is Lipschitz continuous for $\tau_j, j = 1, 2$, where $\tau = (\tau_1, \tau_2) \in \mathbb{D}$; (iii) there exists constants $\rho_b > 0$ and $k_b > 0$ such that $\inf_{\mathbf{b}_j \in \mathcal{B}(\rho_b)} \text{eigmin} \mathbf{A}_j(\mathbf{b}_j) > k_b$, where $\mathcal{B}(\rho_b) = \{\mathbf{b}_j \in R^p : \inf_{\tau \in \mathbb{D}} \|\mathbf{b}_j - \beta_{j0}(\tau_j)\| \leq \rho_b\}$, for $j = 1, 2, \boldsymbol{\tau} = (\tau_1, \tau_2)$; (iv) the copula function is differentiable with continuous partial derivatives with regard to τ_1 and τ_2 for any \mathbf{Z} .
- C4. (i) $\sup_{\boldsymbol{\tau} \in \mathbb{D}} \|\boldsymbol{\gamma}_0(\boldsymbol{\tau})\|$ is bounded above; (ii) there exists a constant $k_r > 0$ such that $\inf_{\boldsymbol{\tau} \in \mathbb{D}} \text{eigmin}\{\mathbf{J}(\boldsymbol{\gamma}_0(\boldsymbol{\tau}); \boldsymbol{\tau})\} > k_r$.
- C5. (i) For $j = 1, 2$, $f_j(t|\mathbf{z})$ are continuously differentiable with bounded derivatives; (ii) for $j = 1, 2$, $\partial h_j(t_1, t_2|\mathbf{Z})/\partial t_j$ are continuously differentiable with bounded derivatives.

Remarks: Condition C1 assumes the boundedness of covariates, which is often met in practice. Condition C2 is satisfied in many clinical settings with administrative censoring. Conditions C3 (i)-(iii) assume uniform boundedness of marginal densities and smoothness of coefficient processes, which are standard assumptions for marginal quantile regression methods with independent censoring data, and are usually reasonable in practice. Condition C3 (iv) implies the boundedness of $h_j(t_1, t_2|\mathbf{z})$ in (t_1, t_2) and \mathbf{z} . Condition C4 lists standard assumptions for quantile association models, which include the boundedness of $\boldsymbol{\gamma}_0(\boldsymbol{\tau})$ and the

identifiability of $\gamma_0(\boldsymbol{\tau})$. Condition C5 contains mild assumptions for adopting a consistent covariance estimator.

Let $M_i^{G_j}(s) = I(Y_{ji} \leq s, \delta_j = 0) - \int_0^\infty I(Y_{ji} \geq u) d\Lambda^{G_j}(u)$, where $\Lambda^{G_j}(u)$ is the cumulative hazard function for the censoring variable C_j . Define $\xi_{1,ji}(\tau_j) = \mathbf{Z}_i(I[Y_{ji} \leq g\{\mathbf{Z}_i^T \boldsymbol{\beta}_{j0}(\tau_j)\}])\delta_j G_j(Y_{ji})^{-1} - \tau_j$ and $\xi_{2,ji}(\tau_j) = \int_0^\infty w(\boldsymbol{\beta}_{j0}(\tau_j), s)P(Y_{ji} \geq s)^{-1} dM_i^{G_j}(s)$, where $w\{\boldsymbol{\beta}_{j0}(\tau_j), s\} = E\{\mathbf{Z}I(Y_j \geq t)I[Y_j \leq g_j\{\mathbf{Z}^T \boldsymbol{\beta}_{j0}(\tau_j)\}]\delta_j G_j(Y_j)^{-1}\}$. Let $\xi_{ji}(\tau_j) = \xi_{1,ji}(\tau_j) - \xi_{2,ji}(\tau_j)$. To obtain the explicit form of the influence function, we here assume the univariate censoring mechanism. Let $Y_i^* = \max(Y_{1i}, Y_{2i})$ and $\delta_i^* = 1 - \delta_{1i}\delta_{2i}$. The univariate censoring function $G(\cdot)$ can be estimated from $\{Y_i^*, \delta_i^*\}_{i=1}^n$, and we denote $\hat{G}(\cdot)$ as the consistent estimator of bivariate censoring function. Let $y^*(t) = \Pr(Y^* \geq t)$, $M_i^G(t) = N_i^G(t) - \int_0^\infty I(Y_i^* \geq s) d\Lambda^G(s)$ and $\xi_i^*(\boldsymbol{\tau}) = \int_0^\infty w^*\{\boldsymbol{\beta}_{10}(\tau_1), \boldsymbol{\beta}_{20}(\tau_2), s\}P(Y_i^* \geq s)^{-1} dM_i^G(s)$, where $w^*\{\boldsymbol{\beta}_{10}(\tau_1), \boldsymbol{\beta}_{20}(\tau_2), s\} = E[\mathbf{Z}I(Y^* \geq s)I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\}\delta_1\delta_2 G(Y^*)^{-1}]$.

Theorem 3. Suppose models (3.3) and (3.4) hold for $\tau \in \mathbb{D}$. Under conditions C1-C5, $\sup_{\boldsymbol{\tau} \in \mathbb{D}} \|\hat{\gamma}(\boldsymbol{\tau}) - \gamma_0(\boldsymbol{\tau})\| \xrightarrow{p} 0$.

Theorem 4. Suppose models (3.3) and (3.4) hold for $\tau \in \mathbb{D}$. Under conditions C1-C5, $n^{1/2}\{\hat{\gamma}(\boldsymbol{\tau}) - \gamma_0(\boldsymbol{\tau})\}$ converges weakly to a zero-mean Gaussian process for $\tau \in \mathbb{D}$ with a limiting covariance matrix which equals

$$\boldsymbol{\Omega}(\boldsymbol{\tau}', \boldsymbol{\tau}) = \mathbf{J}\{\gamma_0(\boldsymbol{\tau}'); \boldsymbol{\tau}'\}^{-1} E\{\psi_i(\boldsymbol{\tau}')\psi_i(\boldsymbol{\tau})^T\} \mathbf{J}\{\gamma_0(\boldsymbol{\tau}); \boldsymbol{\tau}\}^{-T}, \quad (3.5)$$

where

$$\begin{aligned} \psi_i(\boldsymbol{\tau}) = & \mathbf{Z}_i \frac{I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}(\tau_2)\} \delta_{1i} \delta_{2i}}{G(Y_{1i}, Y_{2i})} - \mathbf{Z}_i \chi\{\exp(\mathbf{Z}_i^T \gamma_0(\boldsymbol{\tau})); \boldsymbol{\tau}\} \\ & - \xi_i^*(\boldsymbol{\tau}) - \sum_{j=1}^2 \mathbf{P}_j\{\boldsymbol{\beta}_{10}(\tau_1), \boldsymbol{\beta}_{20}(\tau_2)\} \mathbf{A}_j^{-1}\{\boldsymbol{\beta}_{j0}(\tau_j)\} \xi_{ji}(\tau_j). \end{aligned} \quad (3.6)$$

The proofs for Theorems 3 and 4 are detailed in Appendices D and E.

3.2.4 Covariance Estimations

The covariance estimation under quantile regression models is often difficult, because the asymptotic covariance matrix involves unknown conditional density functions due to the unsmoothness of the corresponding estimating equations. In previous studies, several methods, such as using a nonparametric density estimator and resampling, have been considered. These methods, however, have computational problems and poor performance with small sample sizes. In this study, we employ the idea of the induced smoothing procedure that was proposed by [Brown and Wang \(2005\)](#) to estimate the covariance matrices for both marginal regression estimators and the conditional association coefficient estimators. The induced smoothing method smooths the original estimating equation by using a “pseudo-Bayesian” approach. This technique has been well extended to the quantile regression models ([Brown and Wang, 2007](#); [Wang et al., 2009](#); [Pang et al., 2012](#); [Li et al., 2014](#)).

In the following, we first estimate the influence functions for the marginal regression estimators, and then further derive the influence function for the conditional association estimator. Currently the univariate censoring scenario is considered to simplify the asymptotic representation. [Peng and Fine \(2009\)](#) extended the sampling-based technique proposed by [Huang \(2002\)](#) to the estimation of the covariance matrix. Here, we propose to use the induced smoothing technique, which is resampling-free and achieves better finite-sample performances than bootstrap-based’s estimators ([Pang et al., 2012](#)). The similar work can be found in [Pang et al. \(2012\)](#), where they developed an induced smoothing procedure for [Bang and Tsiatis \(2002\)](#)’s estimator.

It has been shown in [Peng and Fine \(2009\)](#) that, under regularity conditions, $S_{nj}(\mathbf{b}_j, \tau_j)$ converges weakly to a mean-zero Gaussian process with covariance $\Sigma_j(\tau'_j, \tau_j) = \text{cov}\{\xi_j(\tau_j)\}$ and the estimators, $\widehat{\beta}_j(\tau_j)$, are consistent with the true values $\beta_{j0}(\tau_j)$. The asymptotic distribution for $n^{1/2}\{\widehat{\beta}_j(\tau_j) - \beta_{j0}(\tau_j)\}$ would be a mean-zero Gaussian process with covariance

$$\mathbf{D}_j(\beta_{j0}; \tau_j) = \mathbf{A}_j\{\beta_{j0}(\tau_j)\}^{-1} \Sigma_j(\tau'_j, \tau_j) \mathbf{A}_j\{\beta_{j0}(\tau_j)\}^{-T},$$

where $\mathbf{A}_j(\mathbf{b}_j) = E[\mathbf{Z}^{\otimes 2} f_j\{g_j(\mathbf{Z}^T \mathbf{b}_j)\}] = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{Z}_i^{\otimes 2} f_j\{g_j(\mathbf{Z}_i^T \mathbf{b}_j)\}$, for $j = 1, 2$. We now adopt the induced smoothing approach to $S_{nj}(\mathbf{b}_j, \tau_j)$ and obtain a consistent estimate

of $\mathbf{A}_j\{\boldsymbol{\beta}_{j0}(\tau_j)\}$. First, by the asymptotic normality of $\widehat{\boldsymbol{\beta}}_j(\tau_j)$, we can approximately write $\widehat{\boldsymbol{\beta}}_j(\tau_j) = \boldsymbol{\beta}_{j0}(\tau_j) + \mathbf{B}_j^{1/2}V_j$, where $\mathbf{B}_j = n^{-1}\mathbf{D}_j$, $V_j \sim N(0, I_p)$, and I_p is the $p \times p$ identity matrix. We can regard $\widehat{\boldsymbol{\beta}}_j(\tau_j)$ as a random perturbation of $\boldsymbol{\beta}_{j0}(\tau_j)$. Hence, we define a considerably smoother estimating function,

$$\begin{aligned}\widetilde{S}_{nj}(\mathbf{b}_j, \mathbf{B}_j; \tau_j) &= E_{V_j}\{S_{nj}(\mathbf{b}_j + \mathbf{B}_j^{1/2}V_j; \tau_j, \mathbf{B}_j)\} \\ &= n^{-1} \sum_{i=1}^n \mathbf{Z}_i \left[\frac{\delta_{ji}}{\widehat{G}_j(Y_{ji})} \Phi \left\{ \frac{\mathbf{Z}_i^T \mathbf{b}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \right\} - \tau_j \right],\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Through the smoothed estimating function $\widetilde{S}_{nj}(\mathbf{b}_j, \mathbf{B}_j; \tau_j)$, the estimator of $\mathbf{A}_j(\mathbf{b}_j)$ can be achieved from the derivative of the smoothed estimating equation with respect to \mathbf{b}_j , which has the form,

$$\widetilde{\mathbf{A}}_j(\mathbf{b}_j, \mathbf{B}_j) = \frac{\partial \widetilde{S}_{nj}(\mathbf{b}_j, \mathbf{B}_j; \tau_j)}{\partial \mathbf{b}_j} = n^{-1} \sum_{i=1}^n \frac{\delta_{ji} \mathbf{Z}_i^{\otimes 2}}{\widehat{G}_j(Y_{ji}) \sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \mathbf{b}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \right\}, \quad (3.7)$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution. Given \mathbf{B}_j , we can obtain the estimator $\widetilde{\boldsymbol{\beta}}_j$ by solving $\widetilde{S}_{nj}(\widetilde{\boldsymbol{\beta}}_j, \mathbf{B}_j; \tau_j) = 0$ and then plug it into (3.7) to get the estimator,

$$\widetilde{\mathbf{A}}_j(\widetilde{\boldsymbol{\beta}}_j, \mathbf{B}_j) = n^{-1} \sum_{i=1}^n \frac{\delta_{ji} \mathbf{Z}_i^{\otimes 2}}{\widehat{G}_j(Y_{ji}) \sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \widetilde{\boldsymbol{\beta}}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \right\}.$$

In general, the matrix \mathbf{B}_j is unknown. Hence, we develop an iterative algorithm to achieve the optimal solutions for both $\boldsymbol{\beta}_{j0}(\tau_j)$ and \mathbf{B}_j . The procedure is given below:

Step 0. Set the initial $\widetilde{\mathbf{B}}_j^{(0)} = n^{-1}I_p$ and $\widetilde{\boldsymbol{\beta}}_j^{(0)} = \widehat{\boldsymbol{\beta}}_j(\tau_j)$, and let $\widehat{\boldsymbol{\Sigma}}_j(\tau'_j, \tau_j) = n^{-1} \sum_{i=1}^n \widehat{\xi}_{ji}^{\otimes 2}$.

Step 1. In the k -th iteration, update $\widetilde{\boldsymbol{\beta}}_j^{(k)}$ by solving $\widetilde{S}_{nj}(\widetilde{\boldsymbol{\beta}}_j^{(k-1)}, \widetilde{\mathbf{B}}_j^{(k-1)}; \tau_j) = 0$.

Step 2. Update $\widetilde{\mathbf{B}}_j^{(k)} = n^{-1}(\widetilde{\mathbf{A}}_j^{(k)})^{-1} \widehat{\boldsymbol{\Sigma}}_j(\tau'_j, \tau_j) (\widetilde{\mathbf{A}}_j^{(k)})^{-T}$, where

$$\widetilde{\mathbf{A}}_j^{(k)} = n^{-1} \sum_{i=1}^n \frac{\delta_{ji} \mathbf{Z}_i^{\otimes 2}}{\widehat{G}_j(Y_{ji}) \sqrt{\mathbf{Z}_i^T \widetilde{\mathbf{B}}_j^{(k-1)} \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \widetilde{\boldsymbol{\beta}}_j^{(k)} - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \widetilde{\mathbf{B}}_j^{(k-1)} \mathbf{Z}_i}} \right\}.$$

Step 3. Repeat Steps 1-2 until convergence.

This algorithm is computationally efficient and leads to a consistent covariance estimator $\tilde{\mathbf{D}}_j = n\tilde{\mathbf{B}}_j^{(k)}$ after the convergence of the iterations. More theoretical justifications and arguments were discussed in Pang et al. (2012).

We next estimate the influence function for $\gamma(\tau)$. In Section 3.2.3, the proposed asymptotic distribution of $\sqrt{n}\{\hat{\gamma}(\tau) - \gamma_0(\tau)\}$ is a mean-zero Gaussian process with covariance $\Omega(\tau', \tau) = \mathbf{J}\{\gamma_0(\tau); \tau\}^{-1} E\{\psi_i(\tau')\psi_i(\tau)^T\} \mathbf{J}\{\gamma_0(\tau); \tau\}^{-T}$. Define a consistent estimator of $\mathbf{J}(\gamma_0(\tau); \tau)$ as

$$\hat{\mathbf{J}}(\hat{\gamma}; \tau) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i^{\otimes 2} \chi' \{\exp(\mathbf{Z}_i^T \hat{\gamma})\} \exp(\mathbf{Z}_i^T \hat{\gamma}),$$

where $\chi'(u) = \partial \chi(u) / \partial u$. To have a consistent estimator for $E\{\psi_i(\tau')\psi_i(\tau)^T\}$, we firstly estimate $\mathbf{P}_j\{\beta_{10}(\tau_1), \beta_{20}(\tau_2)\}$, where,

$$\mathbf{P}_j\{\beta_{10}(\tau_1), \beta_{20}(\tau_2)\} = E(\mathbf{Z}^{\otimes 2} h_j[g_1\{\mathbf{Z}^T \beta_{10}(\tau_1)\}, g_2\{\mathbf{Z}^T \beta_{20}(\tau_2)\}] g'_j\{\mathbf{Z}^T \beta_{j0}(\tau_j)\}).$$

However, estimating $\mathbf{P}_j\{\beta_{10}(\tau_1), \beta_{20}(\tau_2)\}$ directly is difficult since $\mathbf{P}_j(\cdot, \cdot)$ involves an unknown partial density function $h_j(\cdot, \cdot)$. To address this issue, we propose an induced-smoothing type estimator for $\mathbf{P}_j\{\beta_{10}(\tau_1), \beta_{20}(\tau_2)\}$, for $j = 1, 2$. For brevity, we simply the notation, such as $\hat{\beta}_j = \hat{\beta}_j(\tau_j)$ and $\hat{\gamma} = \hat{\gamma}(\tau)$. Adapting the induced smoothing methods for the marginal quantile effects, we obtain a smoothed estimating function, where

$$\begin{aligned} \tilde{\mathbf{W}}_{nj}^{\hat{G}}(\mathbf{b}_j; \hat{\beta}_{j^*}, \hat{\gamma}, \tilde{\mathbf{B}}_j) &= E_{V_j}\{\mathbf{W}_n^{\hat{G}}(\mathbf{b}_j + \tilde{\mathbf{B}}_j^{1/2} V_j, \hat{\beta}_{j^*}, \hat{\gamma}; \tau)\} \\ &= n^{-1} \sum_{i=1}^n \mathbf{Z}_i \left[\frac{\delta_{1i} \delta_{2i} I\{g_{j^*}^{-1}(Y_{j^*i}) \leq \mathbf{Z}_i^T \hat{\beta}_{j^*}\}}{\hat{G}(Y_{1i}, Y_{2i})} \Phi \left\{ \frac{\mathbf{Z}_i^T \mathbf{b}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \right\} - \chi\{\exp(\mathbf{Z}_i^T \hat{\gamma}); \tau\} \right], \end{aligned}$$

where $\tilde{\mathbf{B}}_j$ is the induced-smoothing type estimator for \mathbf{B}_j from marginal quantile models, and $j^* = 3 - j$, for $j = 1, 2$. Therefore, $\mathbf{P}_j\{\beta_{10}(\tau_1), \beta_{20}(\tau_2)\}$ can be estimated by

$$\begin{aligned} \hat{\mathbf{P}}_j^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \tilde{\mathbf{B}}_j) &= \frac{\partial \tilde{\mathbf{W}}_{nj}^{\hat{G}}(\mathbf{b}_j; \hat{\beta}_{j^*}, \hat{\gamma}, \tilde{\mathbf{B}}_j)}{\partial \mathbf{b}_j} \Big|_{\mathbf{b}_j = \hat{\beta}_j} \\ &= n^{-1} \sum_{i=1}^n \frac{\mathbf{Z}_i^{\otimes 2} \delta_{1i} \delta_{2i} I\{g_{j^*}^{-1}(Y_{j^*i}) \leq \mathbf{Z}_i^T \hat{\beta}_{j^*}\}}{\hat{G}(Y_{1i}, Y_{2i}) \sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \hat{\beta}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \right\}, \end{aligned}$$

where $j^* = 3 - j$, for $j = 1, 2$.

Coupled with these estimators, an estimation of the influence function $\psi_i(\boldsymbol{\tau})$ can be obtained by the following procedures:

Step A. For $j = 1, 2$, employ Steps 0-3 in the algorithm to assess $\tilde{\mathbf{B}}_j$.

Step B. For $j = 1, 2$, let $j^* = 3 - j$ and define

$$\hat{\mathbf{P}}_j^{\hat{G}}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \tilde{\mathbf{B}}_j) = n^{-1} \sum_{i=1}^n \frac{\mathbf{Z}_i^{\otimes 2} \delta_{1i} \delta_{2i} I\{g_{j^*}^{-1}(Y_{j^*i}) \leq \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_{j^*}\}}{\hat{G}(Y_{1i}, Y_{2i}) \sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \right\},$$

and

$$\hat{\mathbf{A}}_j(\hat{\boldsymbol{\beta}}_j) = n^{-1} \sum_{i=1}^n \frac{\delta_{ji} \mathbf{Z}_i^{\otimes 2}}{\hat{G}_j(Y_{ji}) \sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \right\}.$$

Step C. Plug in $\hat{\boldsymbol{\beta}}_j$, $\hat{\boldsymbol{\gamma}}$, $\hat{\xi}_{ji}$, $\hat{\xi}_i^*$, and the above estimates into (3.6). The resulting estimator for $\psi_i(\boldsymbol{\tau})$ is

$$\begin{aligned} \hat{\psi}_i(\boldsymbol{\tau}) = & \mathbf{Z}_i \frac{I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_1, g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}_2\} \delta_{1i} \delta_{2i}}{\hat{G}(Y_{1i}, Y_{2i})} - \mathbf{Z}_i \chi\{\exp(\mathbf{Z}_i^T \hat{\boldsymbol{\gamma}}); \boldsymbol{\tau}\} \\ & - \hat{\xi}_i^* - \sum_{j=1}^2 \hat{\mathbf{P}}_j\{\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \tilde{\mathbf{B}}_j\} \hat{\mathbf{A}}_j^{-1}\{\hat{\boldsymbol{\beta}}_j\} \hat{\xi}_{ji}(\tau_j). \end{aligned}$$

By applying Steps A-C, we can further propose an estimator for the covariance matrix,

$$\hat{\boldsymbol{\Omega}}(\boldsymbol{\tau}', \boldsymbol{\tau}) = \hat{\mathbf{J}}\{\hat{\boldsymbol{\gamma}}; \boldsymbol{\tau}'\}^{-1} \left\{ n^{-1} \sum_{i=1}^n \hat{\psi}_i(\boldsymbol{\tau}') \hat{\psi}_i(\boldsymbol{\tau})^T \right\} \hat{\mathbf{J}}\{\hat{\boldsymbol{\gamma}}; \boldsymbol{\tau}\}^{-T},$$

where

$$\hat{\mathbf{J}}(\hat{\boldsymbol{\gamma}}; \boldsymbol{\tau}) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i^{\otimes 2} \chi' \{\exp(\mathbf{Z}_i^T \hat{\boldsymbol{\gamma}})\} \exp(\mathbf{Z}_i^T \hat{\boldsymbol{\gamma}}),$$

which is a consistent estimator of $\mathbf{J}\{\hat{\boldsymbol{\gamma}}; \boldsymbol{\tau}'\}$. The details of the consistency of $\hat{\psi}_i(\boldsymbol{\tau})$ are provided in Appendix F.

3.3 SIMULATION

In this section, numerical simulations are conducted to investigate the finite-sample performance of our proposed models. Without loss of generality, we focus on the case that $\tau_1 = \tau_2 = \tau$. Two covariates are generated, Z_1 and Z_2 , where Z_1 is a standard normal distributed variate and truncated at -2 and 2 , and Z_2 is a Bernoulli distributed variate with $p = 0.5$. Denote $\mathbf{Z} = (Z_1, Z_2)$. For two event times T_1 and T_2 , we generate the marginal quantile regression models with the exponential link function, $g_1(t) = g_2(t) = \exp(t)$, and

$$\log Q_1(\tau|\mathbf{Z}) = 0.2\Phi^{-1}(\tau) + 0.2Z_1 + \{0.4\Phi^{-1}(\tau) - 0.2\Phi^{-1}(\tau)\}Z_2,$$

$$\log Q_2(\tau|\mathbf{Z}) = 0.3\Phi^{-1}(\tau) - 0.2Z_1 + 0.5Z_2,$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the cumulative distribution function of the standard normal distribution. From the above models, the effect of Z_1 is constant across τ for both $\log Q_1(\tau|\mathbf{Z})$ and $\log Q_2(\tau|\mathbf{Z})$. The effect of Z_2 is constant on $\log Q_2(\tau|\mathbf{Z})$ but varies for $\log Q_1(\tau|\mathbf{Z})$ by τ .

To generate the association structure, we consider that (T_1, T_2) follows a flipped-Clayton model with parameter θ when $Z_2 = 1$, and they are conditionally independent when $Z_2 = 0$. Specifically,

$$\mathcal{C}(\boldsymbol{\tau}|Z_2 = 1) = \tau_1 + \tau_2 - 1 + \max[\{(1 - \tau_1)^{-\theta} + (1 - \tau_2)^{-\theta} - 1\}^{-1/\theta}, 0],$$

where $\theta = \exp(1)$. We have that the underlying quantile association models is generated by

$$\log qor(\boldsymbol{\tau}|\mathbf{Z}) = \log[\chi^{-1}\{\mathcal{C}(\boldsymbol{\tau}|\mathbf{Z})\}]Z_2,$$

where $\chi^{-1}(y)$ is the inverse function of χ and is monotone increasing in y . Note that Z_2 is the only covariate that affects the association, and the influence of Z_2 on the association structure is increasing by $\boldsymbol{\tau}$. Under this setting, the true value of $\boldsymbol{\gamma}_0(\boldsymbol{\tau})$ is $(0, 0, \log[\chi^{-1}\{\mathcal{C}(\boldsymbol{\tau}|\mathbf{Z})\}])^T$.

With the generated covariates \mathbf{Z} , we now simulate two event times under the assumed association structure. For subjects with $Z_2 = 1$, the assumed copula function is the flipped-Clayton model, which implies that the bivariate survival function $S(t_1, t_2) = \Pr(T_1 > t_1, T_2 > t_2)$ follows a Clayton model. Hence, we simulate the Clayton event times via a frailty model

with the frailty $W \sim \text{gamma}(1/\theta, 1)$, $\theta > 0$. The corresponding Laplace transformation function for W is denoted as $\eta(t) = E\{\exp(-tW)\} = (1+t)^{-1/\theta}$ with the inverse function $\eta^{-1}(u) = u^{-\theta} - 1$. It is easy to show that the marginal survival functions $S_j(t_j) = 1 - F_j(t_j) = \eta(t_j)$, for $j = 1, 2$, and the bivariate survival function satisfies $S(t_1, t_2) = \eta[\eta^{-1}\{S_1(t_1)\} + \eta^{-1}\{S_2(t_2)\}]$, for subjects with $Z_2 = 1$. That is, we can construct the assumed copula model based on the frailty framework. We begin by generating two variates U_1 and U_2 from $\text{Unif}(0, 1)$, and draw W from the $\text{gamma}(1/\theta, 1)$ distribution. Define $U_j^* = (1 - Z_2)U_j + Z_2[1 - \{1 - \log(U_j)/W\}^{-1/\theta}]$, for $j = 1, 2$. We use U_j^* to generate the event time T_j through its underlying marginal regression model, for $j = 1, 2$. Without loss of generality, we assume the univariate censoring setting, and generate C from a mixture distribution of $\text{Unif}(0, c_b)$ with probability 0.8, and a point mass at c_b with probability 0.2. With this mixture distribution, the regularity condition for the censoring is satisfied. In practice, c_b represents the endpoint for the study. Therefore, the observed bivariate survival data are $(Y_1, Y_2, \delta_1, \delta_2, \mathbf{Z})$, where $Y_j = \min(T_j, C)$ and $\delta_j = I(Y_j \leq C)$, for $j = 1, 2$.

We performed 2000 simulations with sample sizes $n = 200$ or 400 to examine the proposed model and the estimation of covariance. We set $c_b = 6$ or 4 so that the percentage of the censoring is about 20% or 30%. For $\tau = 0.2, 0.25, 0.3, 0.4, 0.5, 0.6$, Tables 3.1 and 3.2 present the results of the empirical bias (Bias), the empirical standard error (empSE), the average estimated standard error (estSE) and the empirical coverage probability of 95% Wald-type confidence intervals (COV) for (I) $\hat{\beta}_1(\tau)$, (II) $\hat{\beta}_2(\tau)$ and (III) $\hat{\gamma}(\tau)$, under different censoring rates. From the top and middle parts of Table 3.1, we can see that, with 20% censoring rate, the estimated marginal quantile coefficients are largely unbiased across all τ s; the induced smoothing standard errors agree well with the empirical ones, and the Wald-type confidence intervals based on the induced smoothing standard errors are close to the nominal level 95%. The results for the conditional association coefficients are shown in Table 3.1 (III) for 20% censoring and in Table 3.2 (III) for 30% censoring. The biases for the association coefficients are noticeably larger than those for the marginal regression coefficients. This is as expected, since the estimation of the association parameters depends on the estimation of marginal effects. Nevertheless, the biases are largely negligible, suggesting that the proposed estimator $\hat{\gamma}(\tau)$ provides accurate estimation of the true association effect across τ s. The

standard errors based on induced smoothing tend to be slightly larger than the empirical standard errors. Consequently, the coverage rates of Wald-type confidence intervals are greater than the nominal level 95%. [Li et al. \(2014\)](#) also reported an inflated coverage rate of the confidence interval that was constructed based on the induced smoothing standard error for uncensored pairs. However, as the sample size increases, we observe that the coverage rates of Wald-type confidence intervals are closer to the nominal level 95%. This result implies that our proposed procedure performs reasonably on the covariance estimation. With 30% censoring rate, the results in Table 3.2 give similar conclusions as under 20% censoring rate, though the estimated standard deviations tend to be larger under 30% censoring than those under 20% censoring. Again the coverage rates are getting closer to the nominal level 95% when the sample size increases.

The estimation of covariate effects on the local association depends on the marginal quantile estimation. To understand how the strength of dependency may be affected by the marginal models, we further simulate a scenario, where the association depends on both covariates, and estimate the covariate effects on the quantile association while deliberately leaving out one important covariate in the marginal models. More specifically, we generate the marginal regression models using the previous setting, and introduce $Z_3 = I(Z_1 > 0)$ to the association model via

$$\begin{aligned} \log qor(\boldsymbol{\tau}|\mathbf{Z}) &= \log[\chi^{-1}\{\mathcal{C}_1(\boldsymbol{\tau}|\mathbf{Z})\}]Z_3 + \log[\chi^{-1}\{\mathcal{C}_2(\boldsymbol{\tau}|\mathbf{Z})\}]Z_2 \\ &\quad + (\log[\chi^{-1}\{\mathcal{C}_3(\boldsymbol{\tau}|\mathbf{Z})\}] - \log[\chi^{-1}\{\mathcal{C}_2(\boldsymbol{\tau}|\mathbf{Z})\}] - \log[\chi^{-1}\{\mathcal{C}_1(\boldsymbol{\tau}|\mathbf{Z})\}])Z_2Z_3 \\ &= \gamma^{(0)}(\boldsymbol{\tau}) + \gamma^{(1)}(\boldsymbol{\tau})Z_3 + \gamma^{(2)}(\boldsymbol{\tau})Z_2 + \gamma^{(3)}(\boldsymbol{\tau})Z_2Z_3, \end{aligned}$$

where

$$\mathcal{C}_i(\boldsymbol{\tau}|\mathbf{Z}) = \tau_1 + \tau_2 - 1 + \max\{[(1 - \tau_1)^{-\theta_i} + (1 - \tau_2)^{-\theta_i} - 1]^{-1/\theta_i}, 0\},$$

with $\theta_1 = \exp(-1)$, $\theta_2 = \exp(0)$ and $\theta_3 = \exp(0.5)$. Under this scenario, the association between two survival times is affected by the values of Z_2 and Z_3 , and the bivariate survival times are independent only when $Z_2 = Z_3 = 0$. Furthermore, the interaction term of Z_2 and Z_3 has a very small effect on the association, as compared to the main effects, Z_2 and Z_3 . In the estimation, we include Z_2 , Z_3 , and Z_2Z_3 in the association model, while excluding Z_1

Table 3.1: Simulation results for (I) marginal quantiles coefficients for the first event, (II) marginal quantiles coefficients for the second event, and (III) covariate effects on the quantile association under 20% censoring

(I) $\hat{\beta}_1(\tau) = (\hat{\beta}_1^{(0)}(\tau), \hat{\beta}_1^{(1)}(\tau), \hat{\beta}_1^{(2)}(\tau))^T$													
n	τ	$\hat{\beta}_1^{(0)}(\tau)$				$\hat{\beta}_1^{(1)}(\tau)$				$\hat{\beta}_1^{(2)}(\tau)$			
		Bias	empSE	estSE	COV	Bias	empSE	estSE	COV	Bias	empSE	estSE	COV
200	0.2	0	0.031	0.031	0.934	-0.001	0.034	0.033	0.933	0.004	0.070	0.066	0.924
	0.25	0	0.029	0.030	0.940	0	0.033	0.032	0.927	0.002	0.067	0.064	0.926
	0.3	0	0.029	0.029	0.944	0	0.031	0.031	0.940	0.001	0.064	0.062	0.929
	0.4	0	0.028	0.028	0.941	0	0.031	0.030	0.939	0	0.062	0.062	0.930
	0.5	0.001	0.028	0.029	0.936	0	0.030	0.031	0.938	-0.001	0.063	0.062	0.942
	0.6	0.001	0.029	0.030	0.946	0	0.031	0.032	0.942	-0.002	0.065	0.065	0.940
400	0.2	0.001	0.021	0.022	0.942	0.001	0.025	0.023	0.923	0.004	0.047	0.047	0.942
	0.25	0.001	0.020	0.021	0.942	0.001	0.023	0.022	0.927	0.002	0.045	0.045	0.942
	0.3	0.001	0.020	0.020	0.944	0	0.022	0.021	0.935	0.003	0.044	0.044	0.938
	0.4	0.001	0.019	0.020	0.942	0	0.021	0.021	0.937	0.002	0.044	0.043	0.944
	0.5	0.001	0.019	0.020	0.944	0	0.022	0.021	0.943	0.001	0.044	0.044	0.940
	0.6	0.001	0.020	0.021	0.949	0	0.023	0.022	0.938	0.002	0.046	0.046	0.948
(II) $\hat{\beta}_2(\tau) = (\hat{\beta}_2^{(0)}(\tau), \hat{\beta}_2^{(1)}(\tau), \hat{\beta}_2^{(2)}(\tau))^T$													
n	τ	$\hat{\beta}_2^{(0)}(\tau)$				$\hat{\beta}_2^{(1)}(\tau)$				$\hat{\beta}_2^{(2)}(\tau)$			
		Bias	empSE	estSE	COV	Bias	empSE	estSE	COV	Bias	empSE	estSE	COV
200	0.2	0.001	0.045	0.045	0.929	0	0.038	0.037	0.932	0.001	0.066	0.066	0.933
	0.25	0.001	0.044	0.043	0.936	0.001	0.036	0.036	0.935	0	0.063	0.063	0.941
	0.3	0.001	0.042	0.042	0.933	0.001	0.036	0.036	0.936	0	0.062	0.062	0.939
	0.4	0.002	0.041	0.042	0.940	0.001	0.036	0.035	0.931	-0.001	0.061	0.061	0.951
	0.5	0.002	0.042	0.042	0.936	0.002	0.036	0.036	0.938	-0.002	0.062	0.062	0.938
	0.6	0.003	0.044	0.044	0.942	0.001	0.039	0.038	0.940	-0.002	0.066	0.066	0.943
400	0.2	0	0.032	0.032	0.938	0	0.027	0.026	0.935	0.004	0.046	0.046	0.936
	0.25	0	0.030	0.031	0.936	0	0.026	0.025	0.930	0.003	0.044	0.045	0.946
	0.3	0	0.030	0.030	0.946	0	0.026	0.025	0.933	0.003	0.043	0.044	0.946
	0.4	0.001	0.029	0.029	0.934	0	0.025	0.025	0.932	0.002	0.043	0.043	0.944
	0.5	0.001	0.030	0.030	0.941	0	0.026	0.025	0.934	0.002	0.044	0.044	0.947
	0.6	0.001	0.030	0.031	0.949	0	0.027	0.026	0.934	0.001	0.045	0.046	0.945
(III) $\hat{\gamma}(\tau) = (\hat{\gamma}^{(0)}(\tau), \hat{\gamma}^{(1)}(\tau), \hat{\gamma}^{(2)}(\tau))^T$													
n	τ	$\hat{\gamma}^{(0)}(\tau)$				$\hat{\gamma}^{(1)}(\tau)$				$\hat{\gamma}^{(2)}(\tau)$			
		Bias	empSE	estSE	COV	Bias	empSE	estSE	COV	Bias	empSE	estSE	COV
200	0.2	0.014	0.722	0.797	0.976	-0.010	0.664	0.650	0.990	0.040	1.009	1.089	0.977
	0.25	0.053	0.588	0.645	0.972	-0.008	0.554	0.557	0.984	-0.005	0.891	0.934	0.970
	0.3	0.058	0.550	0.570	0.970	-0.013	0.511	0.508	0.970	-0.001	0.836	0.867	0.966
	0.4	0.096	0.462	0.499	0.967	-0.010	0.469	0.466	0.961	-0.023	0.805	0.830	0.964
	0.5	0.109	0.463	0.493	0.965	-0.001	0.472	0.474	0.958	-0.005	0.842	0.879	0.964
	0.6	0.095	0.526	0.556	0.962	-0.001	0.541	0.548	0.966	0.053	1.011	1.034	0.964
400	0.2	-0.028	0.509	0.527	0.978	-0.016	0.422	0.426	0.973	0.072	0.711	0.720	0.964
	0.25	-0.009	0.431	0.441	0.964	-0.014	0.378	0.376	0.958	0.057	0.627	0.634	0.962
	0.3	0.014	0.386	0.390	0.957	-0.016	0.340	0.345	0.962	0.034	0.577	0.590	0.959
	0.4	0.027	0.340	0.344	0.954	-0.007	0.316	0.318	0.952	0.028	0.561	0.566	0.956
	0.5	0.040	0.329	0.341	0.958	-0.008	0.322	0.324	0.958	0.014	0.585	0.595	0.960
	0.6	0.039	0.356	0.381	0.966	-0.008	0.372	0.370	0.954	0.038	0.665	0.683	0.962

Table 3.2: Simulation results for (I) marginal quantiles coefficients for the first event, (II) marginal quantiles coefficients for the second event, and (III) covariate effects on the quantile association under 30% censoring

(I) $\hat{\beta}_1(\tau) = (\hat{\beta}_1^{(0)}(\tau), \hat{\beta}_1^{(1)}(\tau), \hat{\beta}_1^{(2)}(\tau))^T$													
n	τ	$\hat{\beta}_1^{(0)}(\tau)$				$\hat{\beta}_1^{(1)}(\tau)$				$\hat{\beta}_1^{(2)}(\tau)$			
		Bias	empSE	estSE	COV	Bias	empSE	estSE	COV	Bias	empSE	estSE	COV
200	0.2	0	0.032	0.032	0.937	-0.001	0.035	0.034	0.936	0.004	0.072	0.069	0.924
	0.25	0.001	0.030	0.031	0.934	0	0.034	0.033	0.939	0.002	0.068	0.066	0.935
	0.3	0	0.030	0.030	0.942	0	0.033	0.032	0.936	0	0.066	0.065	0.930
	0.4	0.001	0.029	0.030	0.944	0	0.032	0.032	0.940	0	0.064	0.065	0.939
	0.5	0.001	0.029	0.031	0.944	0.001	0.032	0.033	0.946	-0.001	0.066	0.066	0.938
	0.6	0.002	0.031	0.033	0.956	0.001	0.034	0.035	0.948	-0.002	0.069	0.070	0.946
400	0.2	0.001	0.022	0.022	0.944	0.001	0.025	0.024	0.920	0.003	0.049	0.048	0.937
	0.25	0.001	0.021	0.022	0.936	0	0.024	0.023	0.923	0.002	0.047	0.047	0.942
	0.3	0.001	0.021	0.021	0.941	0	0.023	0.022	0.931	0.003	0.046	0.046	0.942
	0.4	0.001	0.020	0.021	0.941	0	0.023	0.022	0.938	0.002	0.046	0.045	0.950
	0.5	0.001	0.021	0.021	0.948	0	0.023	0.023	0.938	0.001	0.047	0.046	0.942
	0.6	0.001	0.021	0.023	0.946	0	0.024	0.024	0.936	0.002	0.049	0.049	0.944
(II) $\hat{\beta}_2(\tau) = (\hat{\beta}_2^{(0)}(\tau), \hat{\beta}_2^{(1)}(\tau), \hat{\beta}_2^{(2)}(\tau))^T$													
n	τ	$\hat{\beta}_2^{(0)}(\tau)$				$\hat{\beta}_2^{(1)}(\tau)$				$\hat{\beta}_2^{(2)}(\tau)$			
		Bias	empSE	estSE	COV	Bias	empSE	estSE	COV	Bias	empSE	estSE	COV
200	0.2	0.002	0.047	0.047	0.932	0.001	0.039	0.039	0.939	0	0.069	0.069	0.938
	0.25	0.002	0.045	0.045	0.932	0.001	0.038	0.038	0.938	0	0.066	0.067	0.945
	0.3	0.002	0.044	0.044	0.942	0.002	0.037	0.038	0.938	-0.001	0.065	0.066	0.942
	0.4	0.003	0.042	0.044	0.948	0.002	0.039	0.038	0.942	-0.001	0.064	0.066	0.951
	0.5	0.003	0.043	0.045	0.942	0.001	0.040	0.039	0.944	-0.001	0.066	0.068	0.948
	0.6	0.003	0.046	0.048	0.946	0.001	0.043	0.043	0.946	-0.001	0.073	0.074	0.944
400	0.2	0	0.032	0.033	0.948	0	0.028	0.027	0.932	0.004	0.048	0.048	0.946
	0.25	0	0.031	0.032	0.946	0	0.027	0.027	0.932	0.002	0.046	0.047	0.946
	0.3	0	0.030	0.031	0.946	0	0.027	0.026	0.930	0.002	0.045	0.046	0.947
	0.4	0.001	0.030	0.031	0.942	0	0.027	0.026	0.933	0.002	0.045	0.046	0.952
	0.5	0.001	0.031	0.032	0.946	0	0.028	0.027	0.936	0.002	0.047	0.048	0.956
	0.6	0.001	0.032	0.033	0.950	0	0.029	0.030	0.938	0.001	0.050	0.052	0.950
(III) $\hat{\gamma}(\tau) = (\hat{\gamma}^{(0)}(\tau), \hat{\gamma}^{(1)}(\tau), \hat{\gamma}^{(2)}(\tau))^T$													
n	τ	$\hat{\gamma}^{(0)}(\tau)$				$\hat{\gamma}^{(1)}(\tau)$				$\hat{\gamma}^{(2)}(\tau)$			
		Bias	empSE	estSE	COV	Bias	empSE	estSE	COV	Bias	empSE	estSE	COV
200	0.2	0.023	0.746	0.835	0.976	0.009	0.703	0.694	0.992	0.032	1.068	1.155	0.968
	0.25	0.061	0.635	0.681	0.970	-0.006	0.608	0.602	0.982	0.004	0.950	1.004	0.964
	0.3	0.059	0.573	0.600	0.969	-0.024	0.550	0.551	0.980	0.023	0.913	0.935	0.972
	0.4	0.101	0.491	0.531	0.966	-0.003	0.515	0.513	0.964	0.019	0.899	0.914	0.962
	0.5	0.115	0.499	0.534	0.962	-0.012	0.540	0.533	0.960	0.039	0.971	0.999	0.968
	0.6	0.103	0.568	0.621	0.968	-0.002	0.619	0.630	0.978	0.060	1.169	1.206	0.967
400	0.2	-0.037	0.531	0.552	0.976	-0.017	0.449	0.454	0.973	0.086	0.740	0.761	0.966
	0.25	-0.011	0.448	0.461	0.974	-0.014	0.410	0.403	0.967	0.074	0.667	0.675	0.963
	0.3	0.013	0.401	0.410	0.956	-0.013	0.374	0.371	0.966	0.047	0.623	0.631	0.956
	0.4	0.030	0.363	0.365	0.951	-0.010	0.349	0.347	0.958	0.039	0.609	0.615	0.956
	0.5	0.043	0.349	0.368	0.957	-0.003	0.359	0.360	0.956	0.031	0.651	0.661	0.960
	0.6	0.050	0.383	0.424	0.968	-0.004	0.426	0.421	0.960	0.054	0.769	0.783	0.965

(Misspecified Marginal Models), or including both Z_1 and Z_2 (Corrected Marginal Models) the marginal model fitting. Sample sizes $n = 400$ and 600 with censoring rate 20% or 50% are considered. The results are given in Tables 3.3, 3.4, 3.5 and 3.6, which include the empirical bias (Bias), the empirical standard error (empSE), the average estimated standard error (estSE) and the empirical coverage probability (COV), as well as the empirical rejection rate (ERR) of testing if $\gamma^{(j)}(\tau) = 0$, $j = 0, 1, 2, 3$. The results show that, when ignoring Z_1 in the marginal models, the association model tends to have biased estimates for the coefficient of Z_3 , $\gamma^{(1)}(\tau)$, and to have a more significant interaction effect, $\gamma^{(3)}$, which should be negligible. From this discovery, we suggest that practitioners should include all important potential risk factors in the marginal models to avoid spurious association effects due to residual effects of covariates on the marginals.

3.4 DATA ANALYSIS

We illustrate our proposed method by using an age-related macular degeneration (AMD) data from the Age-Related Eye Disease Study (ARDES-Group, 1999), which was designed to assess the risk factors for the development and the progression of AMD. This cohort study has collected data on several risk factors at baseline and times to progression of AMD in both eyes. We want to explore the risk factors for the association between AMD progression times in two eyes using the proposed quantile-based association model, while adjusting for covariate effects on the marginals. Data from 630 Caucasian patients who had at least one eye in moderated AMD stage at baseline are used in the current analysis. The bivariate survival times are the time to progression of AMD in the left and right eyes. Three potential risk factors, age, smoking status (Never, Former and Current), and the baseline AMD severity score (SevereBL-L or SevereBL-R), are considered in the marginal model for AMD progression in the left or right eye. For the conditional association model, instead including both eyes AMD severity scores, we adopt the average of AMD severity scores (AvgSevereBL) in both eyes at baseline to avoid the collinearity issue. The censoring rates for the left and right eyes are 47% and 44%, respectively. Since each bivariate survival pair is

Table 3.3: Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 20% censoring rate, n=400, and 2000 runs

		Marginal models without Z_1 (Misspecified Marginal Models)						Marginal models with both Z_1 and Z_2 (Corrected Marginal Models)					
τ	$\gamma^{(0)}(\tau)$	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0	-0.765	-0.765	0.682	0.939	1.000	0.000	-0.014	-0.014	0.737	0.816	0.983	0.017
0.25	0	-0.693	-0.693	0.684	0.770	0.971	0.029	0.015	0.015	0.645	0.696	0.983	0.017
0.3	0	-0.606	-0.606	0.625	0.668	0.904	0.096	0.037	0.037	0.600	0.635	0.973	0.027
0.4	0	-0.477	-0.477	0.567	0.599	0.892	0.108	0.036	0.036	0.573	0.596	0.964	0.036
0.5	0	-0.363	-0.363	0.604	0.636	0.932	0.068	0.021	0.021	0.614	0.635	0.968	0.032
0.6	0	-0.270	-0.270	0.765	0.795	0.993	0.007	0.017	0.017	0.753	0.778	0.985	0.015

		(Misspecified Marginal Models)						(Corrected Marginal Models)					
τ	$\gamma^{(1)}(\tau)$	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.400	-0.322	-0.722	0.938	1.445	0.995	0.000	0.490	0.089	1.105	1.219	0.989	0.025
0.25	0.427	-0.409	-0.836	1.015	1.207	0.953	0.014	0.497	0.070	1.012	1.073	0.977	0.048
0.3	0.456	-0.459	-0.915	0.987	1.056	0.898	0.042	0.498	0.042	0.958	1.007	0.975	0.056
0.4	0.525	-0.530	-1.055	0.895	0.956	0.841	0.060	0.575	0.050	0.959	0.992	0.960	0.076
0.5	0.612	-0.659	-1.271	0.995	1.049	0.789	0.069	0.662	0.050	1.054	1.087	0.963	0.079
0.6	0.725	-0.870	-1.595	1.390	1.477	0.871	0.019	0.764	0.038	1.305	1.331	0.976	0.059

		(Misspecified Marginal Models)						(Corrected Marginal Models)					
τ	$\gamma^{(2)}(\tau)$	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.916	0.546	-0.370	1.016	1.301	0.990	0.012	0.969	0.052	1.088	1.194	0.984	0.071
0.25	0.981	0.567	-0.414	0.975	1.078	0.961	0.037	1.026	0.045	0.981	1.068	0.980	0.114
0.3	1.050	0.564	-0.485	0.893	0.953	0.942	0.064	1.082	0.033	0.961	1.015	0.974	0.147
0.4	1.204	0.541	-0.663	0.824	0.881	0.895	0.069	1.258	0.054	0.964	1.016	0.971	0.194
0.5	1.386	0.541	-0.846	0.910	0.943	0.856	0.067	1.481	0.095	1.085	1.124	0.969	0.228
0.6	1.609	0.552	-1.057	1.137	1.160	0.852	0.050	1.734	0.124	1.314	1.378	0.980	0.179

		(Misspecified Marginal Models)						(Corrected Marginal Models)					
τ	$\gamma^{(3)}(\tau)$	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.008	1.075	1.067	1.494	1.963	0.976	0.024	-0.132	-0.139	1.719	1.879	0.971	0.028
0.25	0.012	1.225	1.213	1.484	1.687	0.939	0.062	-0.104	-0.116	1.639	1.747	0.970	0.030
0.3	0.014	1.289	1.275	1.436	1.536	0.899	0.101	-0.066	-0.080	1.623	1.709	0.969	0.031
0.4	0.009	1.443	1.433	1.382	1.478	0.862	0.141	-0.091	-0.100	1.704	1.788	0.967	0.032
0.5	-0.011	1.641	1.652	1.562	1.633	0.853	0.145	-0.103	-0.091	1.963	2.036	0.972	0.028
0.6	-0.056	1.950	2.005	2.012	2.125	0.891	0.104	-0.132	-0.076	2.393	2.523	0.984	0.015

Table 3.4: Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 20% censoring rate, n=600, and 2000 runs

Marginal models without Z_1 (Misspecified Marginal Models)								Marginal models with both Z_1 and Z_2 (Corrected Marginal Models)					
τ	$\gamma^{(0)}(\tau)$	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0	-0.951	-0.951	0.678	0.803	0.976	0.024	-0.080	-0.080	0.614	0.655	0.987	0.013
0.25	0	-0.805	-0.805	0.581	0.627	0.839	0.161	-0.038	-0.038	0.521	0.558	0.976	0.024
0.3	0	-0.690	-0.690	0.505	0.539	0.784	0.216	-0.010	-0.010	0.472	0.509	0.974	0.026
0.4	0	-0.512	-0.512	0.462	0.480	0.830	0.170	0.012	0.012	0.463	0.479	0.966	0.034
0.5	0	-0.397	-0.397	0.495	0.511	0.892	0.108	0.019	0.019	0.494	0.509	0.968	0.032
0.6	0	-0.270	-0.270	0.607	0.629	0.974	0.026	-0.003	-0.003	0.597	0.618	0.971	0.029

(Misspecified Marginal Models)								(Corrected Marginal Models)					
τ	$\gamma^{(1)}(\tau)$	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.400	-0.364	-0.764	0.966	1.253	0.973	0.002	0.506	0.106	0.901	0.971	0.978	0.042
0.25	0.427	-0.427	-0.854	0.887	0.993	0.912	0.025	0.504	0.077	0.818	0.860	0.966	0.067
0.3	0.456	-0.466	-0.922	0.806	0.855	0.840	0.056	0.513	0.056	0.783	0.809	0.962	0.084
0.4	0.525	-0.534	-1.060	0.757	0.771	0.722	0.095	0.562	0.037	0.781	0.798	0.962	0.094
0.5	0.612	-0.640	-1.252	0.838	0.850	0.688	0.108	0.642	0.030	0.862	0.876	0.962	0.102
0.6	0.725	-0.873	-1.598	1.138	1.171	0.745	0.066	0.774	0.049	1.028	1.064	0.970	0.089

(Misspecified Marginal Models)								(Corrected Marginal Models)					
τ	$\gamma^{(2)}(\tau)$	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.916	0.662	-0.254	0.952	1.086	0.974	0.039	1.001	0.084	0.920	0.955	0.970	0.142
0.25	0.981	0.622	-0.359	0.815	0.872	0.952	0.068	1.028	0.047	0.816	0.854	0.966	0.189
0.3	1.050	0.602	-0.447	0.723	0.768	0.924	0.092	1.089	0.039	0.760	0.811	0.968	0.238
0.4	1.204	0.554	-0.650	0.684	0.708	0.856	0.110	1.246	0.042	0.794	0.813	0.962	0.320
0.5	1.386	0.541	-0.846	0.741	0.758	0.798	0.104	1.440	0.053	0.886	0.896	0.963	0.350
0.6	1.609	0.530	-1.080	0.910	0.919	0.766	0.072	1.693	0.084	1.082	1.085	0.963	0.326

(Misspecified Marginal Models)								(Corrected Marginal Models)					
τ	$\gamma^{(3)}(\tau)$	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.008	1.210	1.203	1.380	1.646	0.936	0.064	-0.108	-0.116	1.451	1.500	0.962	0.038
0.25	0.012	1.307	1.295	1.303	1.370	0.871	0.132	-0.065	-0.077	1.350	1.396	0.958	0.042
0.3	0.014	1.343	1.329	1.184	1.239	0.838	0.162	-0.067	-0.081	1.319	1.366	0.962	0.038
0.4	0.009	1.448	1.439	1.167	1.189	0.780	0.221	-0.056	-0.066	1.425	1.431	0.956	0.043
0.5	-0.011	1.637	1.648	1.309	1.314	0.762	0.234	-0.067	-0.055	1.616	1.622	0.960	0.040
0.6	-0.056	1.939	1.994	1.630	1.682	0.798	0.191	-0.134	-0.078	1.952	1.983	0.970	0.031

Table 3.5: Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 50% censoring rate, n=400, and 2000 runs

Marginal models without Z_1 (Misspecified Model)								Marginal models with both Z_1 and Z_2 (Corrected Model)					
τ	$\gamma^{(0)}(\tau)$	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0	-0.735	-0.735	0.695	1.012	1.000	0.000	0.003	0.003	0.810	0.906	0.980	0.020
0.25	0	-0.687	-0.687	0.700	0.843	0.993	0.007	0.022	0.022	0.730	0.785	0.986	0.014
0.3	0	-0.610	-0.610	0.648	0.739	0.948	0.052	0.046	0.046	0.684	0.722	0.982	0.018
0.35	0	-0.535	-0.535	0.630	0.690	0.917	0.083	0.056	0.056	0.661	0.694	0.978	0.022
0.4	0	-0.472	-0.472	0.629	0.678	0.911	0.089	0.048	0.048	0.661	0.694	0.976	0.024
0.45	0	-0.412	-0.412	0.661	0.698	0.924	0.076	0.061	0.061	0.701	0.721	0.972	0.028

(Misspecified Model)								(Corrected Model)					
τ	$\gamma^{(1)}(\tau)$	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.400	-0.269	-0.669	0.922	1.543	0.997	0.000	0.498	0.098	1.183	1.364	0.992	0.015
0.25	0.427	-0.351	-0.778	1.018	1.312	0.978	0.005	0.507	0.080	1.098	1.220	0.983	0.025
0.3	0.456	-0.398	-0.855	0.985	1.164	0.942	0.019	0.532	0.076	1.056	1.156	0.977	0.037
0.35	0.489	-0.433	-0.922	0.953	1.088	0.904	0.029	0.567	0.078	1.064	1.140	0.976	0.044
0.4	0.525	-0.475	-1.000	0.978	1.076	0.880	0.045	0.617	0.092	1.101	1.166	0.977	0.054
0.45	0.566	-0.521	-1.087	1.070	1.117	0.851	0.051	0.655	0.089	1.174	1.228	0.976	0.052

(Misspecified Model)								(Corrected Model)					
τ	$\gamma^{(2)}(\tau)$	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.916	0.567	-0.349	1.083	1.448	0.993	0.013	0.949	0.033	1.309	1.414	0.986	0.043
0.25	0.981	0.586	-0.395	1.050	1.232	0.980	0.029	1.048	0.067	1.215	1.286	0.982	0.073
0.3	1.050	0.589	-0.460	0.984	1.106	0.959	0.042	1.104	0.054	1.201	1.243	0.970	0.089
0.35	1.124	0.562	-0.562	0.963	1.057	0.937	0.046	1.163	0.039	1.204	1.247	0.970	0.096
0.4	1.204	0.552	-0.652	0.987	1.056	0.919	0.059	1.278	0.074	1.230	1.297	0.973	0.101
0.45	1.291	0.530	-0.761	1.071	1.101	0.896	0.054	1.369	0.078	1.322	1.388	0.972	0.095

(Misspecified Model)								(Corrected Model)					
τ	$\gamma^{(3)}(\tau)$	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.008	0.964	0.957	1.624	2.180	0.985	0.015	-0.124	-0.131	1.971	2.187	0.985	0.015
0.25	0.012	1.116	1.104	1.627	1.918	0.963	0.038	-0.154	-0.166	1.863	2.058	0.979	0.020
0.3	0.014	1.195	1.181	1.577	1.775	0.939	0.062	-0.131	-0.145	1.901	2.046	0.977	0.023
0.35	0.013	1.296	1.283	1.574	1.734	0.928	0.074	-0.131	-0.144	1.958	2.091	0.977	0.022
0.4	0.009	1.390	1.381	1.663	1.771	0.912	0.091	-0.142	-0.151	2.079	2.223	0.980	0.020
0.45	0.001	1.499	1.498	1.819	1.877	0.898	0.102	-0.139	-0.140	2.262	2.414	0.982	0.017

Table 3.6: Simulation results for covariate effects on the quantile association when the marginal models are misspecified with 50% censoring rate, n=600, and 2000 runs

Marginal models without Z_1 (Misspecified Model)								Marginal models with both Z_1 and Z_2 (Corrected Model)					
τ	$\gamma^{(0)}(\tau)$	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(0)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0	-0.882	-0.882	0.696	0.860	0.999	0.001	-0.055	-0.055	0.698	0.732	0.983	0.017
0.25	0	-0.776	-0.776	0.622	0.690	0.905	0.095	-0.011	-0.011	0.607	0.629	0.975	0.025
0.3	0	-0.668	-0.668	0.561	0.600	0.847	0.153	0.000	0.000	0.554	0.579	0.980	0.020
0.35	0	-0.580	-0.580	0.533	0.557	0.849	0.151	0.009	0.009	0.540	0.559	0.968	0.032
0.4	0	-0.529	-0.529	0.529	0.547	0.852	0.148	0.022	0.022	0.545	0.559	0.967	0.033
0.45	0	-0.467	-0.467	0.549	0.563	0.874	0.126	0.027	0.027	0.557	0.578	0.969	0.031

(Misspecified Model)								(Corrected Model)					
τ	$\gamma^{(1)}(\tau)$	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(1)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.400	-0.367	-0.767	0.966	1.338	0.982	0.005	0.518	0.118	1.009	1.090	0.979	0.041
0.25	0.427	-0.427	-0.853	0.963	1.089	0.930	0.017	0.497	0.070	0.929	0.972	0.966	0.052
0.3	0.456	-0.449	-0.905	0.882	0.948	0.871	0.049	0.513	0.057	0.880	0.922	0.964	0.066
0.35	0.489	-0.467	-0.956	0.843	0.881	0.828	0.065	0.544	0.055	0.877	0.912	0.962	0.072
0.4	0.525	-0.479	-1.005	0.835	0.869	0.809	0.069	0.568	0.042	0.911	0.932	0.962	0.084
0.45	0.566	-0.526	-1.092	0.885	0.902	0.785	0.078	0.613	0.048	0.949	0.981	0.962	0.082

(Misspecified Model)								(Corrected Model)					
τ	$\gamma^{(2)}(\tau)$	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(2)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.916	0.614	-0.302	1.034	1.211	0.976	0.023	1.000	0.084	1.088	1.125	0.973	0.113
0.25	0.981	0.625	-0.355	0.940	0.999	0.962	0.059	1.023	0.043	0.982	1.020	0.969	0.134
0.3	1.050	0.598	-0.451	0.859	0.894	0.933	0.072	1.108	0.058	0.947	0.986	0.971	0.166
0.35	1.124	0.580	-0.544	0.811	0.850	0.917	0.080	1.170	0.046	0.945	0.988	0.970	0.176
0.4	1.204	0.601	-0.603	0.822	0.851	0.899	0.089	1.249	0.045	0.990	1.023	0.971	0.187
0.45	1.291	0.580	-0.711	0.862	0.885	0.872	0.085	1.340	0.049	1.026	1.087	0.973	0.182

(Misspecified Model)								(Corrected Model)					
τ	$\gamma^{(3)}(\tau)$	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR	$\hat{\gamma}^{(3)}(\tau)$	Bias	empSE	estSE	COV	ERR
0.2	0.008	1.134	1.126	1.546	1.826	0.957	0.044	-0.161	-0.169	1.678	1.731	0.969	0.031
0.25	0.012	1.222	1.210	1.476	1.558	0.913	0.089	-0.113	-0.125	1.569	1.627	0.962	0.038
0.3	0.014	1.271	1.257	1.377	1.429	0.877	0.126	-0.096	-0.111	1.556	1.615	0.963	0.037
0.35	0.013	1.330	1.317	1.319	1.391	0.861	0.143	-0.080	-0.094	1.593	1.657	0.962	0.038
0.4	0.009	1.369	1.360	1.366	1.418	0.857	0.144	-0.064	-0.073	1.686	1.748	0.964	0.036
0.45	0.001	1.482	1.481	1.456	1.499	0.855	0.145	-0.068	-0.070	1.815	1.888	0.971	0.029

from the same patient, it is reasonable to assume the univariate censoring on the estimation of the conditional association effects. The overall censoring rate for the univariate censoring is 56%. Due to the heavy censoring, the quantile levels are restricted from 0.12 to 0.4.

The results for the marginal eyes are given in Figures 3.1 and 3.2. For marginal models, age and baseline severity score are mostly significant across quantiles but the smoking status is only significant in a small range. The results for the conditional association model are shown in Figure 3.3. It suggests that only the average AMD severity score has a significantly positive effect on the association at quantile level between 0.15 and 0.35. Moreover, the estimated coefficient for the average AMD severity score at baseline is gradually decreasing but still positive, when the quantile increases. First, the positive coefficient means that the odds of developing AMD in one eye given the developed AMD in the other eye is increasing, when the average AMD severity score at baseline increases. The impact of the average AMD severity score at baseline on the odds ratio is higher at short survival times (low quantiles) than at long survival times (high quantiles) while conditioning on age and smoking status. Thus, the dynamic effects of the average AMD severity score on the local association can be captured by the proposed association model. This result recommends that, for a person who has a large average AMD severity score at baseline, if she/he suffers the development of the AMD in one eye in a short period, it is of importance to monitor the other eye as soon as possible.

3.5 DISCUSSION

In this work, we propose a quantile-based regression model for the association between two event times with independent right censoring. Our proposed method allows dynamic covariate effects on the local association at different quantiles of event times. To have an explicit form of the asymptotic distribution, we assume the univariate censoring when evaluating the bivariate censoring function. In fact, the asymptotic distribution can still be established without the univariate censoring assumption. However it will not lead to a nice equation for the influence function, which has a consistent induced-smoothing type estimator.

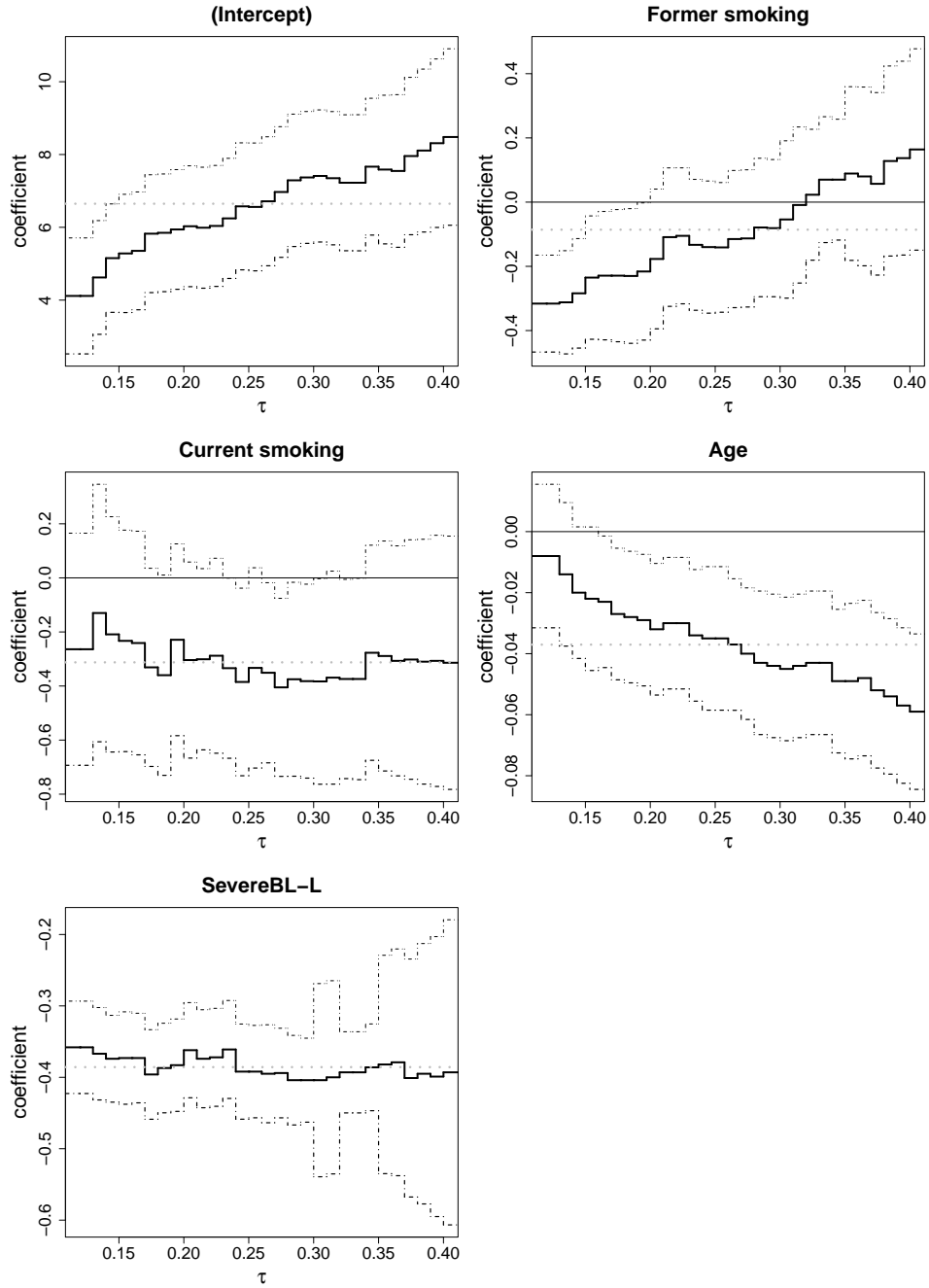


Figure 3.1: The estimated covariate effects for the left eye quantile. The solid bold line is the estimated effect at each quantile level and the dotted line is the average effect across all levels. The dash-dot line is the 95% pointwise confidence interval.

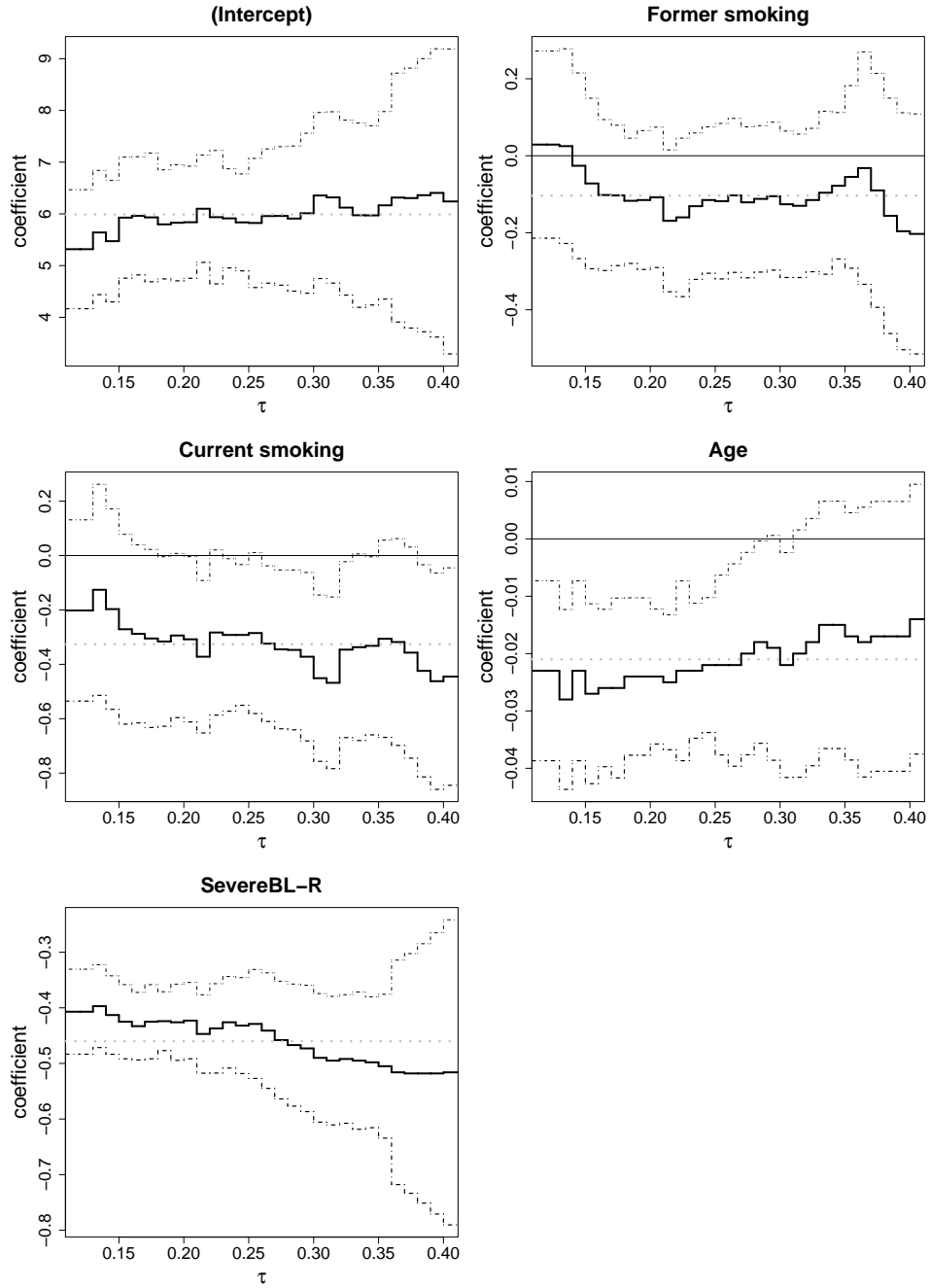


Figure 3.2: The estimated covariate effects for the right eye quantile. The solid bold line is the estimated effect at each quantile level and the dotted line is the average effect across all levels. The dash-dot line is the 95% pointwise confidence interval.

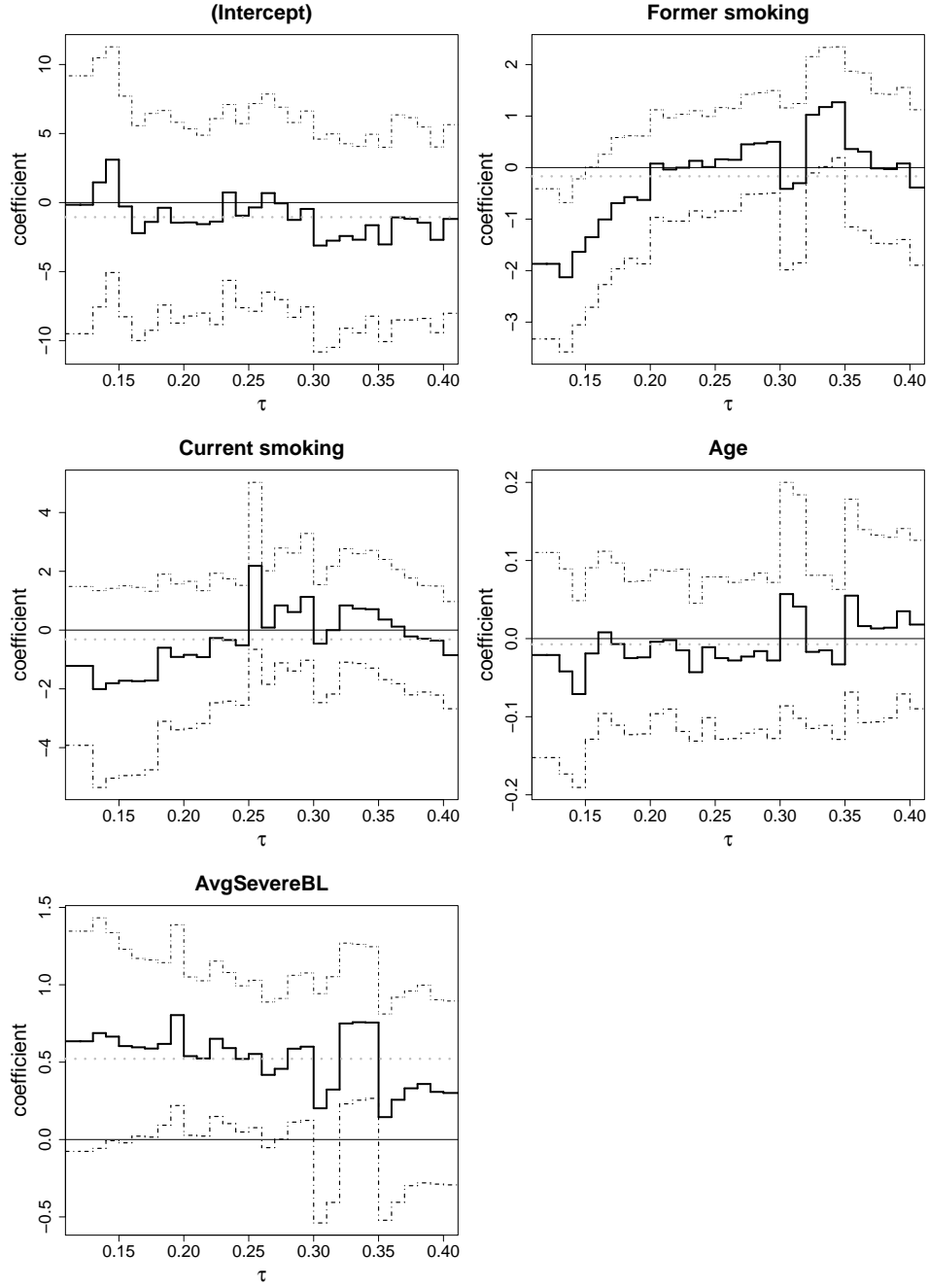


Figure 3.3: The estimated covariate effects on the association using the proposed model. The solid bold line is the estimated effect at each quantile level and the dotted line is the average across all levels. The dash-dot line is the 95% pointwise confidence interval.

With the bivariate survival censoring, the covariance estimation can be achieved by using the bootstrap technique. However, it may result in a larger estimated standard deviation for the quantiled type estimator, which is a common issue in the quantile approaches.

As [Li et al. \(2014\)](#) mentioned, the recommended range of quantiles for a study is associated with its sample size and the number of covariates. In our study which is for the bivariate survival data, the censoring rate also affects the range of quantiles, especially the upper bound level. Thus, the quantile association effects in bivariate survival data may not be estimated at large quantile levels, which are close to the proportion of observed events, where the sample size is small and the number of covariates is large. The restriction, in fact, is much acceptable and universal in any quantile regression analyses of censored data.

4.0 REMARKS AND FUTURE WORKS

In this dissertation, we have proposed two sets of regression models under different types of survival data. One is to examine covariate effects on dynamic treatment regimens for competing risk outcomes. The other focuses on capturing dynamic covariate effects on the local association between two event times in the presence of right censoring.

In the first topic, we propose different regression models for the CIF under a two-stage SMART design. The proposed regression models help practitioners to select an optimal DTR for a patient by taking into account their individual characteristics. Via the comparison of the estimated coefficients in the DTR, practitioners would further appreciate which covariates seem important for the occurrence of the target event. Recently, the SMART design and the DTR have been drawing attention in health care and personalized medicine. Our proposed methods would provide a ready-to-be-used tool for analyzing SMART data with competing risks outcomes.

Our current method extends the existing regression models by properly weighting subjects who are consistent with the DTR of interest. In Chapter 2, we have considered the extensions of two popular models, Fine and Gray’s model (Sections 2.2.2, 2.2.3 and 2.2.4), and Scheike’s model (Section 2.2.5). The proposed weighting method can be easily extended to other regression frameworks for the CIF, such as the multinomial logistic model (Gerds et al., 2012). Corresponding weight functions can be similarly added to the score functions as used in the original model. It would also be of interest to consider double robust estimation (Tsiatis, 2007) in the future to further improve the efficiency of our proposed models.

For the second topic, the proposed quantile-based association regression model for bivariate survival data enables the evaluation of the strength of the local dependency between different quantiles of marginal survival times. More specifically, we use the idea of the copula

to connect with the quantile-based local association qor , and estimate the coefficients for the association at different quantile levels. Our proposed model is very flexible, since it does not require any assumptions on the marginal distributions, and the form of the copula does not need to be specified neither. We examine covariate effects on the quantile association directly while adjusting for risk factors in the marginal distributions. The estimated coefficients can be easily interpreted via the qor .

We notice that the strength of association in two event times may be affected by some residual covariate effects that have not been properly taken into account in the marginal models. For now we recommend considering all potential risk factors when evaluating the marginal distributions. In the future, we plan to further study the connection between the misspecified marginals and the association.

Finally, the dynamic association measurement is useful in capturing the local dependency. However, it would be desirable if we can connect our quantile association model with some commonly used global association measures, such as Kendall's tau. Some weighted local association across quantile levels may be considered. This will be a topic of future work.

APPENDIX A

THE PROOF OF THEOREM 1

Denote $S^{tw(p)}(\beta, u) = n^{-1} \sum_{i=1}^n w_i^{tw}(u) \tilde{Y}_i^{*tw}(u) \mathbf{X}_i^{\otimes p} \exp(\mathbf{X}_i^T \beta)$, $p = 0, 1, 2$, where

$$S^{tw(1)}(\beta, u) = \partial S^{tw(0)}(\beta, u) / \partial \beta \text{ and } S^{tw(2)}(\beta, u) = \partial^2 S^{tw(0)}(\beta, u) / \partial \beta \partial \beta^T.$$

Let $\bar{\mathbf{X}}^{tw}(\beta, u) = S^{tw(1)}(\beta, u) / S^{tw(0)}(\beta, u)$. We replace $\hat{Q}_{A_1 B_k B'_l, i}(t)$ with $Q_{A_1 B_k B'_l, i}(t)$ in the time-varying weighted score function, and recast this score function in terms of martingale integration, under the true β_0 . Under Fine and Gray's model, the time-varying weighted score function for $A_1 B_k B'_l$ regimen with $Q_{A_1 B_k B'_l, i}(t)$ is

$$U_{A_1 B_k B'_l}^{tw}(\beta) = \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta, u) \} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u),$$

where $w_i^{tw}(u) = \{r_i(t) \tilde{G}^{tw}(u)\} / \{\tilde{G}^{tw}(V_i \wedge u)\}$ and $\tilde{N}_{1i}^{tw}(t) = Q_{A_1 B_k B'_l, i}(t) N_{1i}(t)$. The $r_i(t)$ is the vital status at time t for subject i and $\tilde{G}^{tw}(t)$ is a consistent estimate of the censoring function at time t by using the data with time-varying weights. We apply the Taylor expansion to the time-varying weighted score function and have that

$$\sqrt{n}(\hat{\beta}^{tw} - \beta_0) \approx \Omega^{-1} \left\{ n^{-1/2} U_{A_1 B_k B'_l}^{tw}(\beta_0) \right\},$$

where

$$\begin{aligned}
\Omega &= \lim_{n \rightarrow \infty} -\frac{1}{n} \frac{\partial U_{A_1 B_k B'_l}^{tw}(\beta)}{\partial \beta} \Big|_{\beta=\beta_0} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^\infty \frac{S^{tw(2)}(\beta_0, u) S^{tw(0)}(\beta_0, u)^T - S^{tw(1)}(\beta_0, u) S^{tw(1)}(\beta_0, u)^T}{S^{tw(0)}(\beta_0, u) S^{tw(0)}(\beta_0, u)^T} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^\infty \left\{ \frac{S^{tw(2)}(\beta_0, u)}{S^{tw(0)}(\beta_0, u)} - \bar{\mathbf{X}}^{tw}(\beta_0, u)^{\otimes 2} \right\} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u).
\end{aligned}$$

We rewrite the time-varying weight as $w_i^{tw}(u) = \tilde{G}^{tw}(u)I(\Delta_i = 1)/\tilde{G}^{tw}(V_i \wedge u) + I(V_i > u)I(\Delta_i = 0)$ and obtain a consistent estimator of Ω ,

$$\hat{\Omega}^{tw} = \frac{1}{n} \sum_{i=1}^n Q_{A_1 B_k B'_l, i}(V_i) \left\{ \frac{S^{tw(2)}(\hat{\beta}^{tw}, V_i)}{S^{tw(0)}(\hat{\beta}^{tw}, V_i)} - \bar{\mathbf{X}}^{tw}(\hat{\beta}^{tw}, V_i)^{\otimes 2} \right\} \Delta_i I(\epsilon_i = 1).$$

In order to derive the asymptotical distribution for $n^{-1/2}U_{A_1 B_k B'_l}^{tw}(\beta_0)$, we reformulate the score function $U_{A_1 B_k B'_l}^{tw}(\beta)$ in terms of counting processes.

Define $\tilde{A}_i^{*tw}(\beta, t) = \int_0^t \tilde{Y}_i^{*tw}(s) \lambda_{10}(s) \exp(\mathbf{X}_i^T \beta) ds$ and

$$\tilde{M}_{1i}^{tw}(t, \beta) = \tilde{N}_{1i}^{tw}(t) - \tilde{A}_i^{*tw}(\beta, t) = \tilde{N}_{1i}^{tw}(t) - \int_0^t \tilde{Y}_i^{*tw}(s) \lambda_{10}(s) \exp(\mathbf{X}_i^T \beta) ds.$$

Then, the time-varying weighted score function is equivalent to

$$U_{A_1 B_k B'_l}^{tw}(\beta) = \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta, u) \} w_i^{tw}(u) d\tilde{M}^{tw}(u, \beta).$$

Let

$$\tilde{\mathbf{X}}^{tw}(\beta, t) = \frac{\frac{1}{n} \sum_{i=1}^n \tilde{w}_i(t) Y_i^{*tw}(t) \mathbf{X}_j \exp(\mathbf{X}_j^T \beta)}{\frac{1}{n} \sum_{i=1}^n \tilde{w}_i(t) Y_i^{*tw}(t) \exp(\mathbf{X}_j^T \beta)}$$

and

$$\tilde{w}_i(u) = \frac{r_i(u)G(u)}{G(V_i \wedge u)}.$$

We demonstrate that

$$\begin{aligned}
n^{-1/2}U_{A_1 B_k B'_l}^{tw}(\beta) &= n^{-1/2} \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta, u) \} w_i^{tw}(u) d\tilde{M}_{1i}^{tw}(u, \beta) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta, u) \} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta) \\
&\quad + n^{-1/2} \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta, u) \} \{ w_i^{tw}(u) - \tilde{w}_i(u) \} d\tilde{M}_{1i}^{tw}(u, \beta). \quad (\text{A.1})
\end{aligned}$$

By the uniform convergence of $\tilde{G}^{tw}(\cdot)$ to $G(\cdot)$, the second term in (A.1) is

$$\begin{aligned}
& n^{-1/2} \sum_{i=1}^n \int_0^\infty \{\mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta, u)\} \{w_i^{tw}(u) - \tilde{w}_i(u)\} d\tilde{M}_{1i}^{tw}(u, \beta) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\infty \{\mathbf{X}_i - \tilde{\mathbf{X}}^{tw}(\beta, u)\} \{w_i^{tw}(u) - \tilde{w}_i(u)\} d\tilde{M}_{1i}^{tw}(u, \beta) \\
&\quad + n^{-1/2} \sum_{i=1}^n \int_0^\infty \{\tilde{\mathbf{X}}^{tw}(\beta, u) - \bar{\mathbf{X}}^{tw}(\beta, u)\} \{w_i^{tw}(u) - \tilde{w}_i(u)\} d\tilde{M}_{1i}^{tw}(u, \beta) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\infty \{\mathbf{X}_i - \tilde{\mathbf{X}}^{tw}(\beta, u)\} \{w_i^{tw}(u) - \tilde{w}_i(u)\} d\tilde{M}_{1i}^{tw}(u, \beta) + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\infty \{\mathbf{X}_i - \tilde{\mathbf{X}}^{tw}(\beta, u)\} \left\{ \frac{\tilde{G}^{tw}(u)}{\tilde{G}^{tw}(V_i \wedge u)} - \frac{G(u)}{G(V_i \wedge u)} \right\} r_i(u) d\tilde{M}_{1i}^{tw}(u, \beta) + o_p(1).
\end{aligned} \tag{A.2}$$

We represent the Kaplan-Meier estimator in a martingale form, which was introduced by Gill (1980), and have that

$$\frac{\tilde{G}^{tw}(u)}{\tilde{G}^{tw}(V_i \wedge u)} - \frac{G(u)}{G(V_i \wedge u)} = -\frac{G(u)I(V_i < u)}{G(V_i)} \sum_{j=1}^n \int_{V_i}^u \frac{d\tilde{M}_j^{c,tw}(s)}{\sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)} + o_p(1),$$

where $\tilde{M}_j^{c,tw}(s) = Q_{A_1 B_k B'_l, j}(s)I(V_j \leq s, \Delta_j = 0) - \int_0^s Q_{A_1 B_k B'_l, j}(t)I(V_j \geq t)d\Lambda^c(t)$ with Λ^c being the cumulative hazard function for the censoring distribution. We can show that $\tilde{M}_j^{c,tw}(s)$ is a martingale by noting that $I(V_j \leq s, \Delta_j = 0) - \int_0^s I(V_j \geq t)d\Lambda^c(t)$ is a martingale with respect to the censoring filtration, $\mathcal{F}_c(s) = \sigma\{I(V_i \leq u, \Delta_i = 0), I(V_i \geq u), \mathbf{X}_i, u \leq s, \epsilon_i, i = 1, \dots, n\}$. Under some regularity conditions, the second term in (A.1) is dominated by the first term in (A.2). Thus,

$$n^{-1/2} U_{A_1 B_k B'_l}^{tw}(\beta) = n^{-1/2} \sum_{i=1}^n \int_0^\infty \{\mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\beta, u)\} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta) + n^{-1/2} \sum_{i=1}^n \mathbf{R}_i(\beta) + o_p(1),$$

where

$$\begin{aligned}
& \sum_{i=1}^n \mathbf{R}_i(\beta) \\
&= - \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \tilde{\mathbf{X}}^{tw}(\beta, u) \} \frac{G(u)I(V_i < u)}{G(V_i)} \left\{ \sum_{j=1}^n \int_{V_i}^u \frac{d\tilde{M}_j^{c,tw}(s)}{\sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)} \right\} \\
&\quad \times r_i(u) d\tilde{M}_{1i}^{tw}(u, \beta) \\
&= - \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty \{ \mathbf{X}_i - \tilde{\mathbf{X}}^{tw}(\beta, u) \} \frac{G(u)I(V_i < u)}{G(V_i)} \left\{ \int_{V_i}^u \frac{d\tilde{M}_j^{c,tw}(s)}{\sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)} \right\} \\
&\quad \times r_i(u) d\tilde{M}_{1i}^{tw}(u, \beta) \\
&= - \sum_{j=1}^n \int_0^\infty \frac{1}{\sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)} \times \\
&\quad \left[\sum_{i=1}^n I(s > V_i) \int_0^\infty \{ \mathbf{X}_i - \tilde{\mathbf{X}}^{tw}(\beta, u) \} I(u \geq s) \frac{G(u)I(V_i < u)}{G(V_i)} r_i(u) d\tilde{M}_{1i}^{tw}(u, \beta) \right] d\tilde{M}_j^{c,tw}(s) \\
&= - \sum_{j=1}^n \int_0^\infty \frac{\left[\sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \tilde{\mathbf{X}}^{tw}(\beta, u) \} \tilde{w}_i(u) I(V_i < s \leq u) d\tilde{M}_{1i}^{tw}(u, \beta) \right]}{\sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)} d\tilde{M}_j^{c,tw}(s).
\end{aligned}$$

Since $S^{tw(p)}(\beta, u)$, $p = 0, 1$, and $r_i(u)/G(V_i \wedge u)$ are not adapted with respect to $\mathcal{F}_1(u) = \sigma\{I(T_i \leq t, \epsilon_i = 1), 1 - I(T_i \leq t, \epsilon_i = 1), \mathbf{X}_i, t \leq u, i = 1, \dots, n\}$, under regularity conditions, we replace both $\bar{\mathbf{X}}^{tw}(\beta, u)$ and $\tilde{\mathbf{X}}^{tw}(\beta, u)$ with

$$\bar{\mathbf{X}}^{tw}(\beta, u) = s^{tw(1)}(\beta, u)/s^{tw(0)}(\beta, u),$$

where

$$s^{tw(p)}(\beta, u) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i^{*tw}(u) \mathbf{X}_j^{\otimes p} \exp(\mathbf{X}_j^T \beta), p = 0, 1, 2.$$

Therefore,

$$\begin{aligned}
& n^{-1/2} U_{A_1 B_k B'_l}^{tw}(\beta_0) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{x}}^{tw}(\beta_0, u) \} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta_0) + n^{-1/2} \sum_{i=1}^n \mathbf{R}_i(\beta_0) + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{x}}^{tw}(\beta_0, u) \} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta_0) \\
&\quad + n^{-1/2} \sum_{i=1}^n \int_0^\infty \frac{-\sum_{j=1}^n \int_0^\infty \{ \mathbf{X}_j - \bar{\mathbf{x}}^{tw}(\beta_0, u) \} \tilde{w}_j(u) I(V_j < s \leq u) d\tilde{M}_{1j}^{tw}(u, \beta_0)}{\sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)} d\tilde{M}_i^{c, tw}(s) \\
&\quad + o_p(1) \\
&= n^{-1/2} \sum_{i=1}^n \left[\int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{x}}^{tw}(\beta_0, u) \} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta_0) + \int_0^\infty \frac{q^{tw}(s, \beta_0)}{\pi^{tw}(s)} d\tilde{M}_i^{c, tw}(s) \right] + o_p(1),
\end{aligned}$$

where

$$q^{tw}(s, \beta_0) = - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \int_0^\infty \{ \mathbf{X}_j - \bar{\mathbf{x}}^{tw}(\beta_0, u) \} \tilde{w}_j(u) I(V_j < s \leq u) d\tilde{M}_{1j}^{tw}(u, \beta_0),$$

and

$$\pi^{tw}(s) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s).$$

Thus, we can get

$$n^{-1/2} U_{A_1 B_k B'_l}^{tw}(\beta_0) = n^{-1/2} \sum_{i=1}^n (\eta_i^{tw} + \psi_i^{tw}) + o_p(1),$$

where

$$\eta_i^{tw} = \int_0^\infty \{ \mathbf{X}_i - \bar{\mathbf{x}}^{tw}(\beta_0, u) \} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta_0),$$

and

$$\psi_i^{tw} = \int_0^\infty \frac{q^{tw}(s, \beta_0)}{\pi^{tw}(s)} d\tilde{M}_i^{c, tw}(s).$$

Therefore, by the multivariate central limit theorem, $n^{-1/2} U_{A_1 B_k B'_l}^{tw}(\beta_0)$ is asymptotically distributed normal with a covariance matrix

$$\Sigma^{tw} = E \{ (\eta_i^{tw} + \psi_i^{tw})(\eta_i^{tw} + \psi_i^{tw})^T \}.$$

The estimate of Σ^{tw} ,

$$\hat{\Sigma}^{tw} = n^{-1} \sum_{i=1}^n (\hat{\eta}_i^{tw} + \hat{\psi}_i^{tw})^{\otimes 2},$$

where

$$\hat{\eta}_i^{tw} = \int_0^\infty \{\mathbf{X}_i - \bar{\mathbf{X}}^{tw}(\hat{\beta}^{tw}, u)\} w_i^{tw}(u) d\hat{M}_{1i}^{tw}(u, \hat{\beta}^{tw}),$$

$$\hat{\psi}_i^{tw} = \int_0^\infty \frac{\hat{q}^{tw}(s, \hat{\beta}^{tw})}{\hat{\pi}^{tw}(s)} d\hat{M}_i^{c,tw}(s),$$

$$\hat{q}^{tw}(s, \hat{\beta}^{tw}) = -\frac{1}{n} \sum_{j=1}^n \int_0^\infty \{\mathbf{X}_j - \bar{\mathbf{X}}^{tw}(\hat{\beta}^{tw}, u)\} w_j(u) d\hat{M}_{1j}^{tw}(u, \hat{\beta}^{tw}) I(V_j < s \leq u),$$

$$\hat{\pi}^{tw}(s) = \frac{1}{n} \sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s),$$

$$w_i^{tw}(u) = \frac{\tilde{G}^{tw}(u)}{\tilde{G}^{tw}(V_i \wedge u)} I(\Delta_i = 1) + I(V_i > u),$$

$$\tilde{G}^{tw}(u) = \prod_{V_j \leq u} \left(1 - \frac{\sum_{i=1}^n Q_{A_1 B_k B'_l, i}(u) I(V_i = u, \Delta_i = 0)}{\sum_{i=1}^n Q_{A_1 B_k B'_l, i}(u) I(V_i \geq u)} \right),$$

$$\begin{aligned} \hat{M}_{1i}^{tw}(u, \hat{\beta}^{tw}) &= Q_{A_1 B_k B'_l, i}(u) I(V_i \leq u, \epsilon_i = 1) \\ &\quad - \int_0^u Q_{A_1 B_k B'_l, i}(t) \{1 - I(V_i < t, \epsilon_i = 1)\} \exp(\mathbf{X}_i^T \hat{\beta}^{tw}) d\hat{\Lambda}_{10}^{tw}(t), \end{aligned}$$

$$\hat{M}_i^{c,tw}(s) = Q_{A_1 B_k B'_l, i}(s) I(V_i \leq s, \epsilon_i = 0) - \int_0^s Q_{A_1 B_k B'_l, i}(t) I(V_i \geq t) d\hat{\Lambda}^{c,tw}(t),$$

$$\hat{\Lambda}_{10}^{tw}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{1}{S^{tw(0)}(\hat{\beta}^{tw}, v)} w_i^{tw}(v) d\tilde{N}_{1i}^{tw}(v),$$

and

$$\hat{\Lambda}^{c,tw}(t) = \int_0^t \frac{\sum_{i=1}^n Q_{A_1 B_k B'_l, i}(v) I(V_i = v, \Delta_i = 0)}{\sum_{i=1}^n Q_{A_1 B_k B'_l, i}(v) I(V_i \geq v)}.$$

Hence, the asymptotical distribution of

$$\sqrt{n}(\hat{\beta}^{tw} - \beta_0) \approx \Omega^{-1} \left\{ n^{-1/2} U_{A_1 B_k B'_l}^{tw}(\beta_0) \right\}$$

is normally distributed with covariance matrix $\Omega^{-1} \Sigma \Omega^{-1}$, and its estimate is $\hat{\Omega}^{tw-1} \hat{\Sigma} \hat{\Omega}^{tw-1}$.

APPENDIX B

THE PROOF OF THEOREM 2

By following the formula in § 2.2, we can estimate the CIF at time t_0 with covariates x_0 ,

$$\hat{F}_{1,A_1 B_k B'_l}^{tw}(t_0; x_0) = 1 - \exp\{-\hat{\Lambda}_1^{tw}(t_0, x_0)\},$$

where $\hat{\Lambda}_1^{tw}(t_0, x_0) = n^{-1} \sum_{i=1}^n \int_0^{t_0} \exp(x_0^T \hat{\beta}^{tw}) \hat{\Lambda}_{10}^{tw}(t)$ is the the cumulative subdistribution hazard function for the cause 1 at time t_0 with covariates x_0 . To construct the limiting distribution of $n^{1/2}\{\hat{F}_{1,A_1 B_k B'_l}^{tw}(t_0; x_0) - F_{1,A_1 B_k B'_l}(t_0; x_0)\}$, we first establish that $n^{1/2}\{\hat{\Lambda}_1^{tw}(t_0, x_0) - \Lambda_1(t_0, x_0)\}$ converges weakly to a Gaussian process on $[0, \tau)$, where $P(X \leq \tau) > 0$. Applying the Taylor expansion around β_0 to $n^{1/2}\hat{\Lambda}_1^{tw}(t_0, x_0)$, we have

$$\begin{aligned} n^{1/2}\hat{\Lambda}_1^{tw}(t_0, x_0) &= n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \hat{\beta}^{tw})}{S^{tw(0)}(\hat{\beta}^{tw}, u)} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u) + n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \left\{ x_0^T \frac{\exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} \right. \\ &\quad \left. - \frac{S^{tw(1)T}(\beta_0, u) \exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)^2} \right\} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u) (\hat{\beta}^{tw} - \beta_0) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u) \\ &\quad + \int_0^{t_0} \{x_0^T - \bar{\mathbf{X}}^{twT}(\beta_0, u)\} \exp(x_0^T \beta_0) \frac{1}{n} \sum_{i=1}^n \frac{w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u)}{S^{tw(0)}(\beta_0, u)} \sqrt{n}(\hat{\beta}^{tw} - \beta_0) + o_p(1) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u) \\ &\quad + \int_0^{t_0} \{x_0^T - \bar{\mathbf{X}}^{twT}(\beta_0, u)\} \exp(x_0^T \beta_0) d\Lambda_{10}(u) \sqrt{n}(\hat{\beta}^{tw} - \beta_0) + o_p(1). \end{aligned}$$

Thus, $n^{1/2}\hat{\Lambda}_1^{tw}(t_0, x_0)$ is asymptotically equivalent to

$$n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} w_i^{tw}(u) d\tilde{N}_{1i}^w(u) + \mathbf{h}^T(t_0, x_0) \sqrt{n}(\hat{\beta}^{tw} - \beta_0), \quad (\text{B.1})$$

where

$$\mathbf{h}(t_0, x_0) = \int_0^{t_0} \{x_0 - \bar{\mathbf{x}}^{tw}(\beta_0, u)\} \exp(x_0^T \beta_0) d\Lambda_{10}(u).$$

Furthermore, we write $n^{1/2}\Lambda_1(t_0, x_0)$ as

$$n^{1/2} \int_0^{t_0} \exp(x_0^T \beta_0) d\Lambda_{10}(u, x_0) = n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{w_i^{tw}(u) Y_i^{tw}(u) \exp(\mathbf{X}_i^T \beta_0)}{S^{tw(0)}(\beta_0, u)} \exp(x_0^T \beta_0) d\Lambda_{10}(u). \quad (\text{B.2})$$

Then, combining (B.1) and (B.2), $n^{1/2}\{\hat{\Lambda}_1^{tw}(t_0, x_0) - \Lambda_1(t_0, x_0)\}$ is asymptotically equivalent to

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} w_i^{tw}(u) d\tilde{N}_{1i}^{tw}(u) + \mathbf{h}^T(t_0, x_0) \sqrt{n}(\hat{\beta}^{tw} - \beta_0) \\ & \quad - n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{w_i^{tw}(u) \tilde{Y}_i^{tw}(u) \exp(\mathbf{X}_i^T \beta_0)}{S^{tw(0)}(\beta_0, u)} \exp(x_0^T \beta_0) d\Lambda_{10}(u) \\ & = n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{S^{tw(0)}(\beta_0, u)} w_i^{tw}(u) d\tilde{M}_{1i}^{tw}(u, \beta_0) + \mathbf{h}^T(t_0, x_0) \sqrt{n}(\hat{\beta}^{tw} - \beta_0), \end{aligned} \quad (\text{B.3})$$

where

$$d\tilde{M}_{1i}^{tw}(u, \beta_0) = d\tilde{N}_{1i}^w(u) - \tilde{Y}_i^{tw}(u) \exp(\mathbf{X}_i^T \beta_0) d\Lambda_{10}(u).$$

We add and subtract $\tilde{w}_i(u)$, use the asymptotically equivalent form of $n^{1/2}(\hat{\beta}^{tw} - \beta_0)$, and replace $S^{tw(0)}(\beta_0, u)$ with $s^{tw(0)}(\beta_0, u)$ to express (B.3) as

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \{\tilde{w}_i(u) + w_i^{tw}(u) - \tilde{w}_i(u)\} d\tilde{M}_{1i}^{tw}(u, \beta_0) \\ & \quad + \mathbf{h}^T(t_0, x_0) \Omega^{-1} n^{-1/2} U_{A_1 B_k B_l'}^{tw}(\beta_0) + o_p(1) \\ & = n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta_0) \\ & \quad + n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \{w_i^{tw}(u) - \tilde{w}_i(u)\} d\tilde{M}_{1i}^{tw}(u, \beta_0) \\ & \quad + \mathbf{h}^T(t_0, x_0) \Omega^{-1} n^{-1/2} \sum_{i=1}^n (\eta_i^{tw} + \psi_i^{tw}) + o_p(1). \end{aligned} \quad (\text{B.4})$$

Similar to the expression in (A.2), the second term in (B.4) can be written as

$$\begin{aligned}
& -n^{-1/2} \sum_{j=1}^n \int_0^\infty \frac{\left\{ \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \tilde{w}_i(u) I(V_i < s \leq u) d\widetilde{M}_{1i}^{tw}(u, \beta_0) \right\}}{\sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s)} d\widetilde{M}_j^{c, tw}(s) + o_p(1) \\
& = n^{-1/2} \sum_{j=1}^n \int_0^\infty \frac{v^{tw}(s, t_0, x_0, \beta_0)}{\pi^{tw}(s)} d\widetilde{M}_j^{c, tw}(s) + o_p(1),
\end{aligned} \tag{B.5}$$

where

$$v^{tw}(s, t_0, x_0, \beta_0) = - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \tilde{w}_i(u) I(V_i < s \leq u) d\widetilde{M}_{1i}^{tw}(u, \beta_0),$$

and

$$\pi^{tw}(s) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n Q_{A_1 B_k B'_l, m}(s) I(V_m \geq s).$$

Based on these results, (B.4) and (B.5), $n^{1/2}\{\hat{\Lambda}_1^{tw}(t_0, x_0) - \Lambda_1(t_0, x_0)\}$ has an asymptotically equivalent form such as

$$\begin{aligned}
& n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \tilde{w}_i(u) d\widetilde{M}_{1i}^{tw}(u, \beta_0) + n^{-1/2} \sum_{j=1}^n \int_0^\infty \frac{v^{tw}(s, t_0, x_0, \beta_0)}{\pi^{tw}(s)} d\widetilde{M}_j^{c, tw}(s) \\
& + \mathbf{h}^T(t_0, x_0) \Omega^{-1} n^{-1/2} \sum_{i=1}^n (\eta_i^{tw} + \psi_i^{tw}) + o_p(1).
\end{aligned} \tag{B.6}$$

The properties of empirical processes show that the first term in (B.6) is tight, and the second term, which is a martingale integral with respect to $v(s, t_0, x_0, \beta_0)/\pi(s)$, is also tight. Finally, the tightness for the third term can be obtained since the nonrandom function $\mathbf{h}^T(t_0, x_0)$ is the only term which is affected by time. Hence, the asymptotically equivalent form for $n^{1/2}\{\hat{\Lambda}_1^{tw}(t_0, x_0) - \Lambda_1(t_0, x_0)\}$, which is a sum of tight functions, converges to a Gaussian process.

Finally, we apply the functional delta method to get the limiting distribution of $n^{1/2}\{\hat{F}_{1, A_1 B_k B'_l}(t_0; x_0) - F_{1, A_1 B_k B'_l}(t_0; x_0)\}$ on the interval $[0, \tau)$. Since

$$\begin{aligned}
& \frac{\partial F_{1, A_1 B_k B'_l}(t_0; x_0)}{\partial \Lambda_1(t_0, x_0)} [n^{1/2}\{\hat{\Lambda}_1^{tw}(t_0, x_0) - \Lambda_1(t_0, x_0)\}] \\
& = \exp(-\Lambda_1(t_0, x_0)) [n^{1/2}\{\hat{\Lambda}_1^{tw}(t_0, x_0) - \Lambda_1(t_0, x_0)\}],
\end{aligned}$$

the $n^{1/2}\{\hat{F}_{1,A_1B_kB'_l}(t_0; x_0) - F_{1,A_1B_kB'_l}(t_0; x_0)\}$ converges weakly to a Gaussian process which has the same limiting distribution as

$$\begin{aligned} \exp\{-\Lambda_1(t_0, x_0)\} \Big\{ & n^{-1/2} \sum_{i=1}^n \int_0^{t_0} \frac{\exp(x_0^T \beta_0)}{s^{tw(0)}(\beta_0, u)} \tilde{w}_i(u) d\tilde{M}_{1i}^{tw}(u, \beta_0) \\ & + n^{-1/2} \sum_{j=1}^n \int_0^\infty \frac{v^{tw}(s, t_0, x_0, \beta_0)}{\pi^{tw}(s)} d\tilde{M}_j^{c,tw}(s) \\ & + \mathbf{h}^T(t_0, x_0) \Omega^{-1} n^{-1/2} \sum_{i=1}^n (\eta_i^{tw} + \psi_i^{tw}) \Big\} + o_p(1). \end{aligned} \quad (\text{B.7})$$

However, it's complicated to evaluate the above form. Hence, we apply the same technique in [Fine and Gray \(1999\)](#) to estimate the variance of $\hat{F}_{1,A_1B_kB'_l}(t_0; x_0)$. First, we create B copies, $\{\mathbf{W}_{bi}, i = 1, \dots, n\}, b = 1, \dots, B$, from a standard normal distribution, and compute $\hat{K}_b^{tw}(t_0, \mathbf{x}_0)$ for $b = 1, \dots, B$, where

$$\begin{aligned} \hat{K}_b^{tw}(t_0, x_0) = & n^{-1/2} \exp\left\{-\hat{\Lambda}^{tw}(t_0, \mathbf{x}_0)\right\} \times \left\{ \sum_{i=1}^n \int_0^{t_0} \frac{\exp(\mathbf{x}_0^T \hat{\beta}^{tw})}{\hat{S}^{tw(0)}(\hat{\beta}^{tw}, u)} w_i(u) d\hat{M}_{1i}^{tw}(u) \mathbf{W}_{bi} \right. \\ & \left. + \sum_{j=1}^n \int_0^\infty \frac{\hat{v}^{tw}(s, t_0, \hat{\beta}^{tw})}{\hat{\pi}^{tw}(s)} d\hat{M}_j^{c,tw}(s, \hat{\beta}^{tw}) \mathbf{W}_{bi} + \hat{\mathbf{h}}^{twT}(t_0, \mathbf{x}_0) \hat{\Omega}^{tw-1} \sum_{i=1}^n (\hat{\eta}_i^{tw} + \hat{\psi}_i^{tw}) \mathbf{W}_{bi} \right\}. \end{aligned}$$

Thus, the standard deviation for the CIF estimator at time t_0 can be estimated by

$$\hat{\sigma}^{tw}(t_0, \mathbf{x}_0) = \sqrt{\frac{1}{nB} \sum_{b=1}^B \hat{K}_b^{tw^2}(t_0, \mathbf{x}_0)}.$$

APPENDIX C

THE ASYMPTOTIC PROPERTIES FOR THE FIXED WEIGHT SCHEIKE MODEL

Under Scheike's model, we first construct the fixed weight estimating equation for the $A_1 B_k B'_l$ regimen. Let $Q_{A_1 B_k B'_l, i} = R_i I(Z_{1i} = k)/\pi_{B_k} + (1 - R_i) I(Z_{2i} = l)/\pi_{B'_l}$, where $\pi_{B_k} = P(Z_1 = k | R = 1)$ and $\pi_{B'_l} = P(Z_2 = l | R = 0)$. Define $\tilde{N}_i^w(t) = Q_{A_1 B_k B'_l, i} N_i(t)$. It can be shown that

$$\begin{aligned}
& E \left[\frac{\Delta_i \tilde{N}_i^w(t)}{G(T_i | \mathbf{X}_i)} \right] \\
&= E \left[E \left\{ \frac{\Delta_i \tilde{N}_i^w(t)}{G(T_i | \mathbf{X}_i)} \middle| T_i, R_i, Z_{1i}, Z_{2i}, \epsilon_i, \mathbf{X}_i \right\} \right] \\
&= E \left[\tilde{N}_i^w(t) \right] \\
&= E \left[Q_{A_1 B_k B'_l, i} N_i(t) \right] \\
&= E \left[E \{ Q_{A_1 B_k B'_l, i} N_i(t) | R_i, T_i, \mathbf{X}_i, \epsilon_i \} \right] \\
&= E \left[N_i(t) E \{ R_i I(Z_{1i} = k)/\pi_{B_k} + (1 - R_i) I(Z_{2i} = l)/\pi_{B'_l} | R_i, T_i, \mathbf{X}_i, \epsilon_i \} \right] \\
&= E \left(N_i(t) \left[\frac{R_i}{\pi_{B_k}} E \{ I(Z_{1i} = k) | R_i, T_i, \mathbf{X}_i, \epsilon_i \} + \frac{(1 - R_i)}{\pi_{B'_l}} E \{ I(Z_{2i} = l) | R_i, T_i, \mathbf{X}_i, \epsilon_i \} \right] \right) \\
&= E \{ N_i(t) \} \\
&= P(T \leq t, \epsilon = 1 | \mathbf{X}, A_1 B_k B'_l) = F_{1, A_1 B_k B'_l}(t; \mathbf{X}).
\end{aligned}$$

Similar to [Scheike et al. \(2008\)](#), we focus on the interval $[a, \tau]$, where $G(\tau | x) > 0$ and $P(T \leq a | x) > 0$. Let $\hat{G}(\cdot)$ be a consistent estimator of $G(\cdot)$ by using the data with fixed

weights, and $\hat{Q}_{A_1 B_k B'_l, i}$ be a consistent estimator of $Q_{A_1 B_k B'_l, i}$. Thus, in general, it can be shown that

$$\begin{aligned} \frac{N\hat{Q}}{\hat{G}} &= \frac{NQ}{G} + \left(\frac{N\hat{Q}}{\hat{G}} - \frac{NQ}{G} \right) = \frac{NQ}{G} + N \left(\frac{G\hat{Q} - \hat{G}Q}{\hat{G}G} \right) \\ &= \frac{NQ}{G} + N \left(\frac{G\hat{Q} - GQ + GQ - \hat{G}Q}{\hat{G}G} \right) \\ &= \frac{NQ}{G} + N \left\{ \frac{G(\hat{Q} - Q) + Q(G - \hat{G})}{\hat{G}G} \right\} = \frac{NQ}{G} + o_p(1). \end{aligned}$$

In our method, we choose to use $\hat{Q}_{A_1 B_k B'_l, i}$ because it provides more information from the data. Let $N_i^w(t) = \hat{Q}_{A_1 B_k B'_l, i} N_i(t)$ and denote $F_{1, A_1 B_k B'_l}(t; \mathbf{X}) \triangleq F_{1i}^*(t; \eta, \gamma)$. The fixed weight estimating equation for the $A_1 B_k B'_l$ regimen, at time t , can be written as $U^*(\eta, \gamma, \hat{G}) = \{U_1^*(\eta, \gamma, \hat{G})(t), U_2^*(\eta, \gamma, \hat{G})\}$ where

$$\begin{aligned} U_1^*(\eta, \gamma, \hat{G})(t) &= \sum_{i=1}^n D_{\eta, i}^T u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta, \gamma) \right\}, \\ U_2^*(\eta, \gamma, \hat{G}) &= \sum_{i=1}^n \int_a^\tau D_{\gamma, i}^T u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta, \gamma) \right\} dt, \end{aligned}$$

and $u_i(t)$ are possibly random weights. To simplify the notation, we write $\eta(t)$ as η in the following demonstration. To solve the equations, a Taylor expansion around (γ_0, η_0) gives

$$U_1^*(\eta, \gamma, \hat{G})(t) \approx U_1^*(\eta_0, \gamma_0, \hat{G})(t) + \sum_{i=1}^n D_{\eta, i}^T u_i(t) D_{\eta, i}(\eta - \eta_0) + \sum_{i=1}^n D_{\gamma, i}^T u_i(t) D_{\gamma, i}(\gamma - \gamma_0) \quad (\text{C.1})$$

and

$$\begin{aligned} U_2^*(\eta, \gamma, \hat{G}) &\approx U_2^*(\eta_0, \gamma_0, \hat{G}) + \left\{ \sum_{i=1}^n \int_a^\tau D_{\gamma, i}^T u_i(t) D_{\gamma, i} dt \right\} (\gamma - \gamma_0) \\ &\quad + \sum_{i=1}^n \int_a^\tau D_{\gamma, i}^T u_i(t) D_{\eta, i}(\eta - \eta_0) dt. \end{aligned} \quad (\text{C.2})$$

Since $U^*(\hat{\eta}, \hat{\gamma}, \hat{G}) = 0$ where $\hat{\eta}$ and $\hat{\gamma}$ are solutions, by (C.1) and (C.2), we have

$$\hat{\eta} - \eta_0 = - \left[\sum_{i=1}^n D_{\eta, i}^T u_i(t) D_{\eta, i} \right]^{-1} \left(U_1^*(\eta_0, \gamma_0, \hat{G})(t) + \sum_{i=1}^n D_{\gamma, i}^T u_i(t) D_{\gamma, i}(\gamma - \gamma_0) \right) \Big|_{\eta=\hat{\eta}, \gamma=\hat{\gamma}} \quad (\text{C.3})$$

and

$$\begin{aligned} & \left\{ \sum_{i=1}^n \int_a^\tau D_{\gamma,i}^T u_i(t) D_{\gamma,i} dt \right\} (\hat{\gamma} - \gamma_0) \\ &= - \left(U_2^*(\eta_0, \gamma_0, \hat{G}) + \sum_{i=1}^n \int_a^\tau D_{\gamma,i}^T u_i(t) D_{\eta,i} (\eta - \eta_0) dt \right) \Big|_{\eta=\hat{\eta}, \gamma=\hat{\gamma}}. \end{aligned} \quad (\text{C.4})$$

Let

$$\begin{aligned} I_\eta(t) &= \sum_{i=1}^n D_{\eta,i}^T u_i(t) D_{\eta,i}, & I_\gamma(t) &= \sum_{i=1}^n D_{\gamma,i}^T u_i(t) D_{\gamma,i}, \\ I_{\eta,\gamma}(t) &= \sum_{i=1}^n D_{\eta,i}^T u_i(t) D_{\gamma,i}, & I_{\gamma,\eta}(t) &= \sum_{i=1}^n D_{\gamma,i}^T u_i(t) D_{\eta,i}, \end{aligned}$$

and

$$C_\gamma = \int_a^\tau \{I_\gamma(t) - I_{\gamma,\eta}^T(t) I_\eta(t)^{-1} I_{\eta,\gamma}(t)\} dt.$$

Combining (C.3) and (C.4) with these notations, we can show that

$$\hat{C}_\gamma(\hat{\gamma} - \gamma_0) = - \left\{ U_2^*(\eta_0, \gamma_0, \hat{G}) - \int_a^\tau \hat{I}_{\gamma,\eta}(t) \hat{I}_\eta^{-1}(t) U_1^*(\eta_0, \gamma_0, \hat{G})(t) dt \right\}, \quad (\text{C.5})$$

where \hat{C}_γ , $\hat{I}_{\gamma,\eta}(t)$ and $\hat{I}_\eta(t)$ are consistent estimators for C_γ , $I_{\gamma,\eta}(t)$ and $I_\eta(t)$. Therefore,

$$\hat{\gamma} - \gamma_0 = -\hat{C}_\gamma^{-1} \int_a^\tau \sum_{i=1}^n \left\{ D_{\gamma,i}^T - \hat{I}_{\gamma,\eta}(t) \hat{I}_\eta^{-1}(t) D_{\eta,i}^T \right\} u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} dt$$

and

$$\hat{\eta} - \eta_0 = -\hat{I}_\eta^{-1}(t) \sum_{i=1}^n D_{\eta,i}^T u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) + D_{\gamma,i}(\hat{\gamma} - \gamma_0) \right\}.$$

Now we further derive the asymptotic distributions of $\sqrt{n}(\hat{\gamma} - \gamma_0)$ and $\sqrt{n}\{\hat{\eta}(t) - \eta_0(t)\}$. First, the vector of fixed coefficients, γ , has a consistent estimator $\hat{\gamma}$ by solving the score functions. Then, we focus on the vector of time-varying coefficients $\eta(t)$. Since $\hat{Q}_{A_1 B_k B'_l, i}$, which is a fixed weight random variable, does not associate with the event time, the regularity conditions and Lemmas A1 and A2 in Scheike et al. (2008) still hold here. Similar to the arguments in Theorem A1 in Scheike et al. (2008), $\hat{\eta}(t)$ is a uniformly consistent estimator for $\eta_0(t)$. For the extension to the time-varying weight Scheike model, $\hat{Q}_{A_1 B_k B'_l, i}(t)$ involves the time to response. It is considerably more complicated than the fixed weight Scheike model. Hence, we do not derive the influence functions for the time-varying Scheike model here. For the fixed weight Scheike model, to establish the influence function for the fixed

and time-varying coefficients, we use Lemma A2 in [Scheike et al. \(2008\)](#), and coupled with (C.3) and (C.5), we have

$$\begin{aligned}
\hat{C}_\gamma(\hat{\gamma} - \gamma_0) &= - \left(U_2^*(\eta_0, \gamma_0, \hat{G}) - \int_a^\tau \hat{I}_{\gamma, \eta}(t) \hat{I}_\eta^{-1}(t) U_1^*(\eta_0, \gamma_0, \hat{G})(t) dt \right) \\
&= - \left[\sum_{i=1}^n \int_a^\tau D_{\gamma, i}^T u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} dt \right. \\
&\quad \left. - \int_a^\tau \hat{I}_{\gamma, \eta}(t) \hat{I}_\eta^{-1} \sum_{i=1}^n D_{\eta, i}^T u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} dt \right] \\
&= - \sum_{i=1}^n \int_a^\tau \left(D_{\gamma, i}^T - \hat{I}_{\gamma, \eta}(t) \hat{I}_\eta^{-1}(t) D_{\eta, i}^T \right) u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} dt \\
&= - \sum_{i=1}^n \int_a^\tau \hat{K}_i(t) u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} dt,
\end{aligned}$$

where $\hat{K}_i(t) = \left(D_{\gamma, i}^T - \hat{I}_{\gamma, \eta}(t) \hat{I}_\eta^{-1}(t) D_{\eta, i}^T \right)$, and

$$\hat{I}_\eta(t)(\hat{\eta} - \eta_0) = - \left[\sum_{i=1}^n D_{\eta, i}^T u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{\hat{G}_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} + \sum_{i=1}^n D_{\eta, i}^T u_i(t) D_{\gamma, i}(\hat{\gamma} - \gamma_0) \right].$$

Let $W_i = \{V_i, \Delta_i \epsilon_i, \mathbf{X}_i, R_i, Z_{1i}, Z_{2i}\}$,

$$\begin{aligned}
A_{\gamma, \eta, G}^1(t, W_i) &= D_{\eta, i}^T u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{G_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} \\
A_{\gamma, \eta, G}^2(W_i) &= \int_a^\tau \hat{K}_i(t) u_i(t) \left\{ \frac{\Delta_i N_i^w(t)}{G_i(T_i | \mathbf{X}_i)} - F_{1i}^*(t; \eta_0, \gamma_0) \right\} dt,
\end{aligned}$$

and

$$A_{\gamma, \eta, G}^3(t, W_i) = A_{\gamma, \eta, G}^1(t, W_i) - D_{\eta, i}^T u_i(t) D_{\gamma, i} \hat{C}_\gamma^{-1} A_{\gamma, \eta, G}^2(W_i).$$

Also, denote

$$\begin{aligned}
\hat{B}_{\gamma, \eta, G}^1(t, W_i) &= -D_{\eta, i}^T u_i(t) \Delta_i N_i^w(t) \left\{ \frac{\hat{G}_i(T_i | \mathbf{X}_i) - G_i(T_i | \mathbf{X}_i)}{\hat{G}_i(T_i | \mathbf{X}_i) G_i(T_i | \mathbf{X}_i)} \right\}, \\
\hat{B}_{\gamma, \eta, G}^2(W_i) &= - \int_a^\tau \hat{K}_i(t) u_i(t) \Delta_i N_i^w(t) \left\{ \frac{\hat{G}_i(T_i | \mathbf{X}_i) - G_i(T_i | \mathbf{X}_i)}{\hat{G}_i(T_i | \mathbf{X}_i) G_i(T_i | \mathbf{X}_i)} \right\} dt,
\end{aligned}$$

and

$$\hat{B}_{\gamma, \eta, G}^3(t, W_i) = \hat{B}_{\gamma, \eta, G}^1(t, W_i) - D_{\eta, i}^T u_i(t) D_{\gamma, i} \hat{C}_\gamma^{-1} \hat{B}_{\gamma, \eta, G}^2(W_i).$$

Hence, by Slutsky's theorem and Lemmas A1 and A2 in [Scheike et al. \(2008\)](#), we have

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = -\sqrt{n} [C_\gamma^{-1} \{A_{\gamma,\eta,G}^2(W) + B_{\gamma,\eta,G}^2(W)\}] + o_p(1),$$

and

$$\sqrt{n}(\hat{\eta}(t) - \eta_0(t)) = -\sqrt{n} [I_\eta^{-1}(t) \{A_{\gamma,\eta,G}^3(W) + B_{\gamma,\eta,G}^3(W)\}] + o_p(1),$$

where $B_{\gamma,\eta,G}^1(t, W) = E(\hat{B}_{\gamma,\eta,G}^1(t, W))$ and $B_{\gamma,\eta,G}^2(W) = E(\hat{B}_{\gamma,\eta,G}^2(W_i))$. Then, the asymptotic covariance matrices of $\sqrt{n}(\hat{\gamma} - \gamma_0)$ and $\sqrt{n}(\hat{\eta}(t) - \eta_0(t))$ can be consistently estimated by

$$\hat{\Sigma}_\gamma = n\hat{C}_\gamma^{-1} \left(\sum_i (\hat{A}_{\gamma,\eta,G}^2(W_i) + \hat{B}_{\gamma,\eta,G}^2(W_i))^{\otimes 2} \right) \hat{C}_\gamma^{-1}$$

and

$$\hat{\Sigma}_\eta = n\hat{I}_\eta^{-1} \left(\sum_i (\hat{A}_{\gamma,\eta,G}^3(W_i) + \hat{B}_{\gamma,\eta,G}^3(W_i))^{\otimes 2} \right) \hat{I}_\eta^{-1}.$$

APPENDIX D

THE PROOF OF THEOREM 3

The proposed estimating equation for $\gamma_0(\tau)$ is defined as

$$\begin{aligned} & \mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \gamma; \tau) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \left(\frac{I[Y_{1i} \leq g_1\{\mathbf{Z}_i^T \hat{\beta}_1(\tau_1)\}, Y_{2i} \leq g_2\{\mathbf{Z}_i^T \hat{\beta}_2(\tau_2)\}]\delta_{1i}\delta_{2i}}{\hat{G}(Y_{1i}, Y_{2i})} - \chi\{\exp(\mathbf{Z}_i^T \gamma); \tau\} \right) = 0. \end{aligned}$$

Let

$$\begin{aligned} & \mathbf{W}_n^G(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \left(\frac{I[Y_{1i} \leq g_1\{\mathbf{Z}_i^T \mathbf{b}_1(\tau_1)\}, Y_{2i} \leq g_2\{\mathbf{Z}_i^T \mathbf{b}_2(\tau_2)\}]\delta_{1i}\delta_{2i}}{G(Y_{1i}, Y_{2i})} - \chi\{\exp(\mathbf{Z}_i^T \gamma); \tau\} \right), \end{aligned}$$

and $\mathbf{W}(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau) = E\{\mathbf{W}_n^G(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau)\}$.

By conditions C2 and the consistency of $\hat{G}(t_1, t_2)$, we have $\sup_{t_1 < k_1, t_2 < k_2} \|\hat{G}(t_1, t_2) - G(t_1, t_2)\| = o(n^{-1/2+r})$, a.s. for every $r > 0$. This result implies that

$$\sup_{\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau \in \mathbb{D}} \|\mathbf{W}_n^{\hat{G}}(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau) - \mathbf{W}_n^G(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau)\| = o_p(1).$$

Let $\mathcal{F} = \{\mathbf{Z}_i(I[Y_{1i} \leq g_1\{\mathbf{Z}_i^T \mathbf{b}_1(\tau_1)\}, Y_{2i} \leq g_2\{\mathbf{Z}_i^T \mathbf{b}_2(\tau_2)\}]\delta_{1i}\delta_{2i}/G(Y_{1i}, Y_{2i}) - \chi\{\exp(\mathbf{Z}_i^T \gamma); \tau\}) : \mathbf{Z}_i \in \mathcal{Z}, \mathbf{b}_1, \mathbf{b}_2, \gamma \in \mathbb{R}^p, \tau \in \mathbb{D}\}$. The class of indicator functions is Donsker, \mathbf{Z}_i and $G(Y_{1i}, Y_{2i})$ are both uniformly bounded, and $G(Y_{1i}, Y_{2i})$ is assumed to be uniformly bounded

away from zero. Hence, the class \mathcal{F} is Donsker which has the Glivenko-Cantelli property. That is,

$$\sup_{\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau \in \mathbb{D}} \|\mathbf{W}_n^G(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau) - \mathbf{W}(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau)\| = o_p(1).$$

Coupled with those results, we have that

$$\sup_{\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau \in \mathbb{D}} \|\mathbf{W}_n^{\hat{G}}(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau) - \mathbf{W}(\mathbf{b}_1, \mathbf{b}_2, \gamma; \tau)\| = o_p(1). \quad (\text{D.1})$$

Notice that the estimating equation procedure and the model's assumptions guarantee that

$$\mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) = 0 \text{ and } \mathbf{W}(\beta_{10}, \beta_{20}, \gamma_0; \tau) = 0.$$

We further have that

$$\begin{aligned} & \mathbf{W}(\beta_{10}, \beta_{20}, \hat{\gamma}; \tau) - \mathbf{W}(\beta_{10}, \beta_{20}, \gamma_0; \tau) \\ &= \mathbf{W}(\beta_{10}, \beta_{20}, \hat{\gamma}; \tau) - \mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) \\ &= \mathbf{W}(\beta_{10}, \beta_{20}, \hat{\gamma}; \tau) - \mathbf{W}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) + \mathbf{W}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) - \mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau). \end{aligned} \quad (\text{D.2})$$

The first two terms in (D.2) can be written as

$$\begin{aligned} & \mathbf{W}(\beta_{10}, \beta_{20}, \hat{\gamma}; \tau) - \mathbf{W}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) \\ &= E\{\mathbf{W}_n(\beta_{10}, \beta_{20}, \hat{\gamma}; \tau)\} - E\{\mathbf{W}_n(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau)\} \\ &= E\{\mathbf{Z}(H[g_1\{\mathbf{Z}^T \beta_{10}(\tau_1)\}, g_2\{\mathbf{Z}^T \beta_{20}(\tau_2)\}|\mathbf{Z}])\} - E\{\mathbf{Z}(H[g_1\{\mathbf{Z}^T \hat{\beta}_1(\tau_1)\}, g_2\{\mathbf{Z}^T \hat{\beta}_2(\tau_2)\}|\mathbf{Z}])\}. \end{aligned} \quad (\text{D.3})$$

We take Taylor's expansion of the last term in (D.3) at $\beta_{j0}(\tau_j)$, $j = 1, 2$, and show that

$$\mathbf{W}(\beta_{10}, \beta_{20}, \hat{\gamma}; \tau) - \mathbf{W}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) = - \sum_{j=1}^2 \mathbf{P}_j\{\beta_{10}(\tau_1), \beta_{20}(\tau_2)\}\{\hat{\beta}_j(\tau_j) - \beta_{j0}(\tau_j)\} + o_p^{\mathbb{D}}(1),$$

where $\mathbf{P}_j(\mathbf{b}_1, \mathbf{b}_2) = E[\mathbf{Z}^{\otimes 2} h_j\{g_1(\mathbf{Z}^T \mathbf{b}_1), g_2(\mathbf{Z}^T \mathbf{b}_2)\}|\mathbf{Z}]$ and $h_j(t_1, t_2|\mathbf{Z}) = \partial H(t_1, t_2|\mathbf{Z})/\partial t_j$. Under conditions C1, C2, and C3(i)-(iii), the $\hat{\beta}_j(\tau_j)$ uniformly converges to $\beta_{j0}(\tau_j)$ (Peng and Fine, 2009). Based on the condition C3(iv), the $\mathbf{P}_j\{\beta_{10}(\tau_1), \beta_{20}(\tau_2)\}$ is uniformly bounded for $\tau = (\tau_1, \tau_2) \in \mathbb{D}$. Therefore, we further have the results that

$$\sup_{\tau \in \mathbb{D}} \|\mathbf{W}(\beta_{10}, \beta_{20}, \hat{\gamma}; \tau) - \mathbf{W}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau)\| = o_p(1). \quad (\text{D.4})$$

With those above results, (D.1), (D.2) and (D.4), it can be shown that

$$\sup_{\boldsymbol{\tau} \in \mathbb{D}} \|\mathbf{W}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \hat{\boldsymbol{\gamma}}; \boldsymbol{\tau}) - \mathbf{W}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \boldsymbol{\gamma}_0; \boldsymbol{\tau})\| = o_p(1). \quad (\text{D.5})$$

Thus, under the condition C4, we use a first-order Taylor's expression of $\mathbf{W}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \hat{\boldsymbol{\gamma}}; \boldsymbol{\tau})$ at $\boldsymbol{\gamma}_0(\boldsymbol{\tau})$ to have that

$$\hat{\boldsymbol{\gamma}}(\boldsymbol{\tau}) - \boldsymbol{\gamma}_0(\boldsymbol{\tau}) = \mathbf{J}(\boldsymbol{\gamma}_0; \boldsymbol{\tau})^{-1} \{\mathbf{W}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \hat{\boldsymbol{\gamma}}; \boldsymbol{\tau}) - \mathbf{W}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \boldsymbol{\gamma}_0; \boldsymbol{\tau})\} + o_p^{\mathbb{D}}(1),$$

where $\mathbf{J}(\boldsymbol{\gamma}; \boldsymbol{\tau}) = \partial E[\mathbf{Z}\chi\{\exp(\mathbf{Z}^T \boldsymbol{\gamma})\}]/\partial \boldsymbol{\gamma} = E[\mathbf{Z}^{\otimes 2} \chi'\{\exp(\mathbf{Z}^T \boldsymbol{\gamma})\} \exp(\mathbf{Z}^T \boldsymbol{\gamma})]$. With the boundedness of $\mathbf{J}(\boldsymbol{\gamma}_0; \boldsymbol{\tau})$ and (D.5), the consistency of $\hat{\boldsymbol{\gamma}}(\boldsymbol{\tau})$ can be obtained, where

$$\sup_{\boldsymbol{\tau} \in \mathbb{D}} \|\hat{\boldsymbol{\gamma}}(\boldsymbol{\tau}) - \boldsymbol{\gamma}_0(\boldsymbol{\tau})\| = o_p(1).$$

APPENDIX E

THE PROOF OF THEOREM 4

To simplify the notation, we first define

$$\mathbf{V}_n^{\hat{G}}(\mathbf{b}_1, \mathbf{b}_2; \boldsymbol{\tau}) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i \frac{I[Y_{1i} \leq g_1\{\mathbf{Z}_i^T \mathbf{b}_1(\tau_1)\}, Y_{2i} \leq g_2\{\mathbf{Z}_i^T \mathbf{b}_2(\tau_2)\}] \delta_{1i} \delta_{2i}}{\hat{G}(Y_{1i}, Y_{2i})},$$

$$\mathbf{V}_n^G(\mathbf{b}_1, \mathbf{b}_2; \boldsymbol{\tau}) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i \frac{I[Y_{1i} \leq g_1\{\mathbf{Z}_i^T \mathbf{b}_1(\tau_1)\}, Y_{2i} \leq g_2\{\mathbf{Z}_i^T \mathbf{b}_2(\tau_2)\}] \delta_{1i} \delta_{2i}}{G(Y_{1i}, Y_{2i})},$$

and $\mathbf{V}(\mathbf{b}_1, \mathbf{b}_2; \boldsymbol{\tau}) = E\{\mathbf{V}_n^G(\mathbf{b}_1, \mathbf{b}_2; \boldsymbol{\tau})\}$.

Lemma 1. *For any sequence $\{\tilde{\boldsymbol{\beta}}_{1n}(\tau_1), \tilde{\boldsymbol{\beta}}_{2n}(\tau_2), \boldsymbol{\tau} = (\tau_1, \tau_2) \in \mathbb{D}\}_{n=1}^\infty$ satisfying $\sup \|\tilde{\boldsymbol{\beta}}_{jn}(\tau_j) - \boldsymbol{\beta}_{j0}(\tau_j)\| = o_p(1)$, for $j = 1, 2$, we have*

$$\sup \|\mathbf{V}_n^G(\tilde{\boldsymbol{\beta}}_{1n}, \tilde{\boldsymbol{\beta}}_{2n}; \boldsymbol{\tau}) - \mathbf{V}_n^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}; \boldsymbol{\tau}) - \mathbf{V}(\tilde{\boldsymbol{\beta}}_{1n}, \tilde{\boldsymbol{\beta}}_{2n}; \boldsymbol{\tau}) + \mathbf{V}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}; \boldsymbol{\tau})\| = o_p(n^{-1/2}).$$

To prove this lemma, we follow the results in ? and the arguments of [Lai and Ying \(1988\)](#). With the boundedness of \mathbf{Z} , it is sufficient to show that

$$\begin{aligned} \sup_{\boldsymbol{\tau} \in \mathbb{D}} \|\text{var}(G(Y_1, Y_2)^{-1} I[Y_1 \leq g_1\{\mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{1n}(\tau_1)\}, Y_2 \leq g_2\{\mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{2n}(\tau_2)\}] \delta_1 \delta_2 \\ - G(Y_1, Y_2)^{-1} I[Y_1 \leq g_1\{\mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1)\}, Y_2 \leq g_2\{\mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\}] \delta_1 \delta_2)\| = o_p(1). \end{aligned} \tag{E.1}$$

The variance form in (E.1) can be written as

$$\begin{aligned}
& \text{var}(G(Y_1, Y_2)^{-1} \delta_1 \delta_2 [I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{1n}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{2n}(\tau_2)\} \\
& \quad - I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\}]) \\
& \leq \mathbb{E}(G(Y_1, Y_2)^{-1} \delta_1 \delta_2 [I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{1n}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{2n}(\tau_2)\} \\
& \quad - I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\}])^2 \\
& = \mathbb{E}|G(Y_1, Y_2)^{-2} \delta_1 \delta_2 [I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{1n}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{2n}(\tau_2)\} \\
& \quad - I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\}])| \\
& \leq \mathbb{E}|G(Y_1, Y_2)^{-2} \delta_1 \delta_2 [I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{1n}(\tau_1)\} - I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1)\}]| \\
& \quad + \mathbb{E}|G(Y_1, Y_2)^{-2} \delta_1 \delta_2 [I\{g_2^{-1}(Y_2) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{2n}(\tau_2)\} - I\{g_2^{-1}(Y_2) \leq \mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\}]| \\
& \leq \mathbb{E}|G(Y_1, Y_2)^{-2} [I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{1n}(\tau_1)\} - I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1)\}]| \\
& \quad + \mathbb{E}|G(Y_1, Y_2)^{-2} [I\{g_2^{-1}(Y_2) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{2n}(\tau_2)\} - I\{g_2^{-1}(Y_2) \leq \mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\}]|.
\end{aligned}$$

With conditions C1 and C3, we have that

$$\mathbb{E}|I\{g_j^{-1}(Y_j) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{jn}(\tau_j)\} - I\{g_j^{-1}(Y_j) \leq \mathbf{Z}^T \boldsymbol{\beta}_{j0}(\tau_j)\}| = o_p^{\mathbb{D}}(1) \text{ for } j = 1, 2$$

Due to the boundedness of $G(t_1, t_2)$ and condition C2, we further can show that

$$\mathbb{E}|G(Y_1, Y_2)^{-2} [I\{g_j^{-1}(Y_j) \leq \mathbf{Z}^T \tilde{\boldsymbol{\beta}}_{jn}(\tau_j)\} - I\{g_j^{-1}(Y_j) \leq \mathbf{Z}^T \boldsymbol{\beta}_{j0}(\tau_j)\}]| = o_p^{\mathbb{D}}(1).$$

This result implies the equation (E.1) which completes the proof of Lemma 1.

Proof of Theorem 4: We began with the proposed estimating equation along with true parameters and formulated that

$$\begin{aligned}
& \mathbf{W}_n^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \boldsymbol{\gamma}_0; \boldsymbol{\tau}) \\
& = \mathbf{W}_n^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \boldsymbol{\gamma}_0; \boldsymbol{\tau}) + \mathbf{V}_n^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}; \boldsymbol{\tau}) - \mathbf{V}_n^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}; \boldsymbol{\tau}) \\
& = \mathbf{W}_n^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \boldsymbol{\gamma}_0; \boldsymbol{\tau}) \\
& \quad + n^{-1} \sum_{i=1}^n \mathbf{Z}_i I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}(\tau_2)\} \delta_{1i} \delta_{2i} \frac{G(Y_{1i}, Y_{2i}) - \hat{G}(Y_{1i}, Y_{2i})}{\hat{G}(Y_{1i}, Y_{2i}) G(Y_{1i}, Y_{2i})}
\end{aligned} \tag{E.2}$$

To express the asymptotic distribution of censoring, we here assume the univariate censoring to obtain the explicit form in equation (E.2). Let $Y_i^* = \max(Y_{1i}, Y_{2i})$ and $\delta_i^* = 1 - \delta_{1i}\delta_{2i}$. The univariate censoring function $G(\cdot)$ can be estimated from $\{Y_i^*, \delta_i^*\}_{i=1}^n$, and we denote $\hat{G}(\cdot)$ as the consistent estimator. Follow the Peng and Fine (2009)'s arguments that $\sup_{t \in [0, v)} \|n^{1/2}\{\hat{G}(t) - G(t)\} - n^{1/2} \sum_{i=1}^n G(t) \int_0^t y^*(s)^{-1} dM_i^G(s)\| \rightarrow 0$, where $y^*(t) = \Pr(Y^* \geq t)$ and $M_i^G(t) = N_i^G(t) - \int_0^\infty I(Y_i^* \geq s) d\Lambda^G(s)$. Along with these results, the last part in equation (E.2) can be approximated by

$$\begin{aligned}
& n^{-1} \sum_{i=1}^n \mathbf{Z}_i I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}(\tau_2)\} \delta_{1i} \delta_{2i} \frac{G(Y_{1i}, Y_{2i}) - \hat{G}(Y_{1i}, Y_{2i})}{\hat{G}(Y_{1i}, Y_{2i}) G(Y_{1i}, Y_{2i})} \\
&= n^{-1} \sum_{i=1}^n \mathbf{Z}_i I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}(\tau_2)\} \delta_{1i} \delta_{2i} \frac{G(Y_i^*) - \hat{G}(Y_i^*)}{\hat{G}(Y_i^*) G(Y_i^*)} \\
&\approx -n^{-3/2} \sum_{i=1}^n \mathbf{Z}_i I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}(\tau_2)\} \delta_{1i} \delta_{2i} \\
&\quad \times \frac{n^{-1/2} \sum_{j=1}^n I(Y_j^* \geq s) y^*(s)^{-1} dM_j^G(s)}{G(Y_i^*)} \\
&\approx -n^{-1} \sum_{i=1}^n \int_0^\infty \left(\sum_{j=1}^n \frac{\mathbf{Z}_j I(Y_j^* \geq s) I\{g_1^{-1}(Y_{1j}) \leq \mathbf{Z}_j^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_{2j}) \leq \mathbf{Z}_j^T \boldsymbol{\beta}_{20}(\tau_2)\} \delta_{1j} \delta_{2j}}{n G(Y_j^*)} \right) \\
&\quad \times \frac{dM_i^G(s)}{y^*(s)} \\
&\approx -n^{-1} \sum_{i=1}^n \sum_{j=1}^n \int_0^\infty w^*\{\boldsymbol{\beta}_{10}(\tau_1), \boldsymbol{\beta}_{20}(\tau_1), s\} \frac{dM_i^G(s)}{y^*(s)} \\
&= -n^{-1} \sum_{i=1}^n \xi_i^*(\boldsymbol{\tau}),
\end{aligned}$$

where $w^*\{\boldsymbol{\beta}_{10}(\tau_1), \boldsymbol{\beta}_{20}(\tau_1), s\} = E[ZI(Y^* \geq s) I\{g_1^{-1}(Y_1) \leq \mathbf{Z}^T \boldsymbol{\beta}_{10}(\tau_1), g_2^{-1}(Y_2) \leq \mathbf{Z}^T \boldsymbol{\beta}_{20}(\tau_2)\} \delta_{1i} \delta_{2i} G(Y^*)^{-1}]$. We then have that

$$\begin{aligned}
\mathbf{W}_n^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \boldsymbol{\gamma}; \boldsymbol{\tau}) &\approx \mathbf{W}_n^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \boldsymbol{\gamma}; \boldsymbol{\tau}) - n^{-1} \sum_{i=1}^n \xi_i^*(\boldsymbol{\tau}) \\
&\approx \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \frac{I[Y_{1i} \leq g_1\{\mathbf{Z}_i^T \boldsymbol{\beta}_{10}(\tau_1)\}, Y_{2i} \leq g_2\{\mathbf{Z}_i^T \boldsymbol{\beta}_{20}(\tau_2)\}]}{G(Y_i^*)} \delta_{1i} \delta_{2i} - \mathbf{Z}_i \chi\{\exp(\mathbf{Z}_i^T \boldsymbol{\gamma}); \boldsymbol{\tau}\} - \xi_i^*(\boldsymbol{\tau}).
\end{aligned} \tag{E.3}$$

With this explicit form, we now further formulate the asymptotical distribution of $n^{1/2}(\hat{\gamma} - \gamma_0)$ via $n^{1/2}\{\mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) - \mathbf{W}_n^G(\beta_{10}, \beta_{20}, \gamma_0; \tau)\}$. Let

$$\mathbf{C}_n(\gamma; \tau) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i \chi\{\exp(\mathbf{Z}_i^T \gamma); \tau\}.$$

We start with showing that

$$\begin{aligned} & n^{1/2}\{\mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) - \mathbf{W}_n^{\hat{G}}(\beta_{10}, \beta_{20}, \gamma_0; \tau)\} \\ &= n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^{\hat{G}}(\beta_{10}, \beta_{20}; \tau)\} - n^{1/2}\{\mathbf{C}_n(\hat{\gamma}; \tau) - \mathbf{C}_n(\gamma_0; \tau)\} \\ &= \text{(I)} - \text{(II)}, \end{aligned}$$

where $\text{(I)} = n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^{\hat{G}}(\beta_{10}, \beta_{20}; \tau)\}$ and $\text{(II)} = n^{1/2}\{\mathbf{C}_n(\hat{\gamma}; \tau) - \mathbf{C}_n(\gamma_0; \tau)\}$. First note that with (I) we have

$$\begin{aligned} & n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^{\hat{G}}(\beta_{10}, \beta_{20}; \tau)\} \\ &= n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau) + \mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau)\} \\ &\quad - n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\beta_{10}, \beta_{20}; \tau) - \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau) + \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau)\} \\ &= n^{1/2}\{\mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau)\} + n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau)\} \\ &\quad - n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\beta_{10}, \beta_{20}; \tau) - \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau)\} \\ &= n^{1/2}\{\mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau)\} + n^{-1/2} \sum_{i=1}^n \mathbf{Z}_i \delta_{1i} \delta_{2i} \left\{ \frac{1}{\hat{G}(Y_i^*)} - \frac{1}{G(Y_i^*)} \right\} \\ &\quad \times [I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \hat{\beta}_1(\tau_1), g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \hat{\beta}_2(\tau_2)\} \\ &\quad - I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \beta_{10}(\tau_1), g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \beta_{20}(\tau_2)\}]. \end{aligned}$$

Since $\sup_i \|\hat{G}^{-1}(Y_i^*) - G^{-1}(Y_i^*)\| = o_p(n^{1/2+r})$, for any $r > 0$, and $E|I\{g_j^{-1}(Y_j) \leq \mathbf{Z}^T \hat{\beta}_j(\tau_j)\} - I\{g_j^{-1}(Y_j) \leq \mathbf{Z}^T \beta_{j0}(\tau_j)\}| = o_p(1)$ for $j = 1, 2$, the equation (I) is dominated by the term of $n^{1/2}\{\mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau)\}$. Coupled with Lemma 1, we have

$$\begin{aligned} \text{(I)} &= n^{1/2}\{\mathbf{V}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^{\hat{G}}(\beta_{10}, \beta_{20}; \tau)\} \\ &\approx n^{1/2}\{\mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau)\} \\ &= n^{1/2}\{\mathbf{V}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}(\beta_{10}, \beta_{20}; \tau)\} + o_p(1). \end{aligned}$$

Using Taylor's expansion of $\mathbf{V}(\hat{\beta}_1, \hat{\beta}_2; \tau)$ at $\beta_{j0}(\tau_j)$, $j = 1, 2$,

$$\mathbf{V}(\hat{\beta}_1, \hat{\beta}_2; \tau) = \mathbf{V}(\beta_{10}, \beta_{20}; \tau) + \sum_{j=1}^2 \frac{\partial \mathbf{V}(\beta_{10}, \beta_{20}; \tau)}{\partial \beta_{j0}} (\hat{\beta}_j - \beta_{j0}) + o_p(n^{-1/2}).$$

Thus,

$$\begin{aligned} n^{1/2} \{ \mathbf{V}_n^G(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}_n^G(\beta_{10}, \beta_{20}; \tau) \} &\approx n^{1/2} \{ \mathbf{V}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}(\beta_{10}, \beta_{20}; \tau) \} \\ &\approx \sum_{j=1}^2 \mathbf{P}_j(\beta_{10}, \beta_{20}) n^{1/2} (\hat{\beta}_j - \beta_{j0}), \end{aligned}$$

where $\mathbf{P}_j(\beta_{10}, \beta_{20}) = \partial \mathbf{V}(\beta_{10}, \beta_{20}; \tau) / \partial \beta_{j0} = E[\mathbf{Z}^{\otimes 2} h_j \{g_1(\mathbf{Z}^T \beta_{10}), g_2(\mathbf{Z}^T \beta_{20})\} | \mathbf{Z}]$. According to [Peng and Fine \(2009\)](#), it has been shown that, under regularity conditions, the estimating equation for β_j , $S_{nj}(\beta_{j0}, \tau_j) = n^{-1} \sum_{i=1}^n \xi_{ji}(\tau_j)$, where $\xi_{ji}(\tau_j) = \xi_{1,ji}(\tau_j) - \xi_{2,ji}(\tau_j)$, $\xi_{1,ji}(\tau_j) = \mathbf{Z}_i(I[Y_{ji} \leq g\{\mathbf{Z}_i^T \beta_{j0}(\tau_j)\}])I(\delta_j = 1)G_j(Y_{ji})^{-1} - \tau_j$ and $\xi_{2,ji}(\tau_j) = \int_0^\infty w(\beta_{j0}(\tau_j), s)P(Y_{ji} \geq s)^{-1}dM_i^{G_j}(s)$. Let $w\{\beta_{j0}(\tau_j), s\} = E(\mathbf{Z}I(Y_j \geq t)I[Y_j \leq g_j\{\mathbf{Z}^T \beta_{j0}(\tau_j)\}])I(\delta_j = 1)G_j(Y_j)^{-1}$ and $M_i^{G_j}(s) = I(Y_{ji} \leq s, \delta_j = 0) - \int_0^\infty I(Y_{ji} \geq u)d\Lambda^{G_j}(u)$, where $\Lambda^{G_j}(u)$ is the cumulative hazard function for the censoring variable C_j . Thus,

$$n^{1/2}(\hat{\beta}_j - \beta_{j0}) \approx -n^{1/2} \mathbf{A}_j(\beta_{j0})^{-1} S_{nj}(\beta_{j0}, \tau_j) = -n^{1/2} \mathbf{A}_j(\beta_{j0})^{-1} \sum_{i=1}^n \xi_{ji}(\tau_j),$$

where $\mathbf{A}_j(\mathbf{b}_j) = E[\mathbf{Z}^{\otimes 2} f_j\{g_j(\mathbf{Z}^T \mathbf{b}_j)\}] = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{Z}_i^{\otimes 2} f_j\{g(\mathbf{Z}_i^T \mathbf{b}_j)\}$, for $j = 1, 2$.

Along with those results, we have that

$$\begin{aligned} \text{(I)} &\approx n^{1/2} \{ \mathbf{V}(\hat{\beta}_1, \hat{\beta}_2; \tau) - \mathbf{V}(\beta_{10}, \beta_{20}; \tau) \} \approx \sum_{j=1}^2 \mathbf{P}_j(\beta_{10}, \beta_{20}) n^{1/2} (\hat{\beta}_j - \beta_{j0}) \\ &\approx -n^{1/2} \sum_{i=1}^n \sum_{j=1}^2 \mathbf{P}_j(\beta_{10}, \beta_{20}) \mathbf{A}_j(\beta_{j0})^{-1} \xi_{ji}(\tau_j). \end{aligned} \tag{E.4}$$

Next, we derive (II), where

$$\text{(II)} = n^{1/2} \{ \mathbf{C}_n(\hat{\gamma}; \tau) - \mathbf{C}_n(\gamma_0; \tau) \} = n^{1/2} \sum_{i=1}^n \mathbf{Z}_i [\chi\{\exp(\mathbf{Z}_i^T \hat{\gamma}); \tau\} - \chi\{\exp(\mathbf{Z}_i^T \gamma_0); \tau\}].$$

We use a first-order Taylor's expression of $n^{1/2} \{ \mathbf{C}_n(\hat{\gamma}; \tau) - \mathbf{C}_n(\gamma_0; \tau) \}$ at $\gamma_0(\tau)$ to have that

$$n^{1/2} \{ \mathbf{C}_n(\hat{\gamma}; \tau) - \mathbf{C}_n(\gamma_0; \tau) \} \approx \hat{\mathbf{J}}(\gamma_0; \tau) n^{1/2} (\hat{\gamma} - \gamma_0) \stackrel{d}{=} \mathbf{J}(\gamma_0; \tau) n^{1/2} (\hat{\gamma} - \gamma_0), \tag{E.5}$$

where $\mathbf{J}(\gamma_0; \tau) = \partial E[\mathbf{Z}\chi\{\exp(\mathbf{Z}^T \gamma_0)\}]/\partial \gamma_0 = E[\mathbf{Z}^{\otimes 2} \chi'\{\exp(\mathbf{Z}^T \gamma_0)\} \exp(\mathbf{Z}^T \gamma_0)]$ has a consistent estimator, $\hat{\mathbf{J}}(\gamma_0; \tau) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i^{\otimes 2} \chi'\{\exp(\mathbf{Z}_i^T \gamma_0); \tau\} \exp(\mathbf{Z}_i^T \gamma_0)$, by the empirical process theory.

Finally, we combine (E.4) and (E.5) to get

$$\begin{aligned} & n^{1/2} \{ \mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) - \mathbf{W}_n^{\hat{G}}(\beta_{10}, \beta_{20}, \gamma_0; \tau) \} \\ & \approx -n^{-1/2} \sum_{i=1}^n \sum_{j=1}^2 \mathbf{P}_j(\beta_{10}, \beta_{20}) \mathbf{A}_j(\beta_{j0})^{-1} \xi_{ji}(\tau_j) - \mathbf{J}(\gamma_0; \tau) n^{1/2} (\hat{\gamma} - \gamma_0). \end{aligned} \quad (\text{E.6})$$

From (E.3), (E.6) and $\mathbf{W}_n^{\hat{G}}(\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}; \tau) = 0$, we derive that

$$\begin{aligned} & \mathbf{J}(\gamma_0; \tau) n^{1/2} (\hat{\gamma} - \gamma_0) \\ & \approx n^{1/2} \mathbf{W}_n^{\hat{G}}(\beta_{10}, \beta_{20}, \gamma_0; \tau) - n^{-1/2} \sum_{i=1}^n \sum_{j=1}^2 \mathbf{P}_j(\beta_{10}, \beta_{20}) \mathbf{A}_j(\beta_{j0})^{-1} \xi_{ji}(\tau_j) \\ & = n^{-1/2} \sum_{i=1}^n \mathbf{Z}_i \frac{I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \beta_{10}, g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \beta_{20}\} \delta_{1i} \delta_{2i}}{G(Y_{1i}, Y_{2i})} - \mathbf{Z}_i \chi\{\exp(\mathbf{Z}_i^T \gamma_0); \tau\} - \xi_i^*(\tau) \\ & \quad - n^{-1/2} \sum_{i=1}^n \sum_{j=1}^2 \mathbf{P}_j(\beta_{10}, \beta_{20}) \mathbf{A}_j(\beta_{j0})^{-1} \xi_{ji}(\tau_j) \\ & = n^{-1/2} \sum_{i=1}^n \left[\mathbf{Z}_i \frac{I\{g_1^{-1}(Y_{1i}) \leq \mathbf{Z}_i^T \beta_{10}, g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \beta_{20}\} \delta_{1i} \delta_{2i}}{G(Y_{1i}, Y_{2i})} - \mathbf{Z}_i \chi\{\exp(\mathbf{Z}_i^T \gamma_0); \tau\} - \xi_i^*(\tau) \right. \\ & \quad \left. - \mathbf{P}_j(\beta_{10}, \beta_{20}) \mathbf{A}_j(\beta_{j0})^{-1} \xi_{ji}(\tau_j) \right] \\ & = n^{-1/2} \sum_{i=1}^n \psi_i(\tau). \end{aligned}$$

Due to the boundedness of $\mathbf{J}(\gamma_0; \tau)^{-1}$, we have that

$$n^{1/2} (\hat{\gamma} - \gamma_0) \approx n^{-1/2} \sum_{i=1}^n \mathbf{J}(\gamma_0; \tau)^{-1} \psi_i(\tau).$$

Under Lipschitz's transformations, the functional class $\mathcal{F}_r = \{\psi_i(\tau) : \tau \in \mathbb{D}\}$ is Donsker. Hence, by applying Donsker's theorem, we can show that $n^{1/2} (\hat{\gamma} - \gamma_0)$ converges weakly to a zero-mean Gaussian process with covariance matrix,

$$\Omega(\tau', \tau) = \mathbf{J}\{\gamma_0(\tau'); \tau'\}^{-1} E\{\psi_i(\tau') \psi_i(\tau)^T\} \mathbf{J}\{\gamma_0(\tau); \tau\}^{-T}.$$

APPENDIX F

JUSTIFICATION FOR THE CONSISTENCY OF THE PROPOSED INFLUENCE FUNCTION

Let $f_j(t|\mathbf{z}) = dF_j(t|\mathbf{z})/dt$, and $h_j(t_1, t_2|\mathbf{Z}) = \partial H(t_1, t_2|\mathbf{Z})/\partial t_j$, for $j = 1, 2$. Denote $\mathbf{A}_j(\mathbf{b}_j) = E[\mathbf{Z}^{\otimes 2} f_j\{g_j(\mathbf{Z}^T \mathbf{b}_j)|\mathbf{Z}\}]$ and $\mathbf{P}_j(\mathbf{b}_1, \mathbf{b}_2) = E[\mathbf{Z}^{\otimes 2} h_j\{g_1(\mathbf{Z}^T \mathbf{b}_1), g_2(\mathbf{Z}^T \mathbf{b}_2)|\mathbf{Z}\} g'_j(\mathbf{Z}^T \mathbf{b}_j)]$, where $g'_j(u) = dg_j(u)/du$. The proposed smoothing estimating functions for β_j and γ are defined as

$$\tilde{S}_{nj}(\mathbf{b}_j, \mathbf{B}_j; \tau_j) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i \left[\frac{\delta_{ji}}{\hat{G}_j(Y_{ji})} \Phi \left\{ \frac{\mathbf{Z}_i^T \mathbf{b}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \right\} - \tau_j \right],$$

and

$$\begin{aligned} & \widetilde{\mathbf{W}}_{nj}^{\hat{G}}(\mathbf{b}_j; \hat{\beta}_{j^*}, \hat{\gamma}, \tilde{\mathbf{B}}_j) \\ &= n^{-1} \sum_{i=1}^n \mathbf{Z}_i \left[\frac{\delta_1 \delta_2 I\{g_{j^*}^{-1}(Y_{j^*i}) \leq \mathbf{Z}_i^T \hat{\beta}_{j^*}\}}{\hat{G}(Y_{1i}, Y_{2i})} \Phi \left\{ \frac{\mathbf{Z}_i^T \mathbf{b}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \tilde{\mathbf{B}}_j \mathbf{Z}_i}} \right\} - \chi\{\exp(\mathbf{Z}_i^T \hat{\gamma}); \boldsymbol{\tau}\} \right]. \end{aligned}$$

Via the smoother estimating equations \tilde{S}_{nj} and $\widetilde{\mathbf{W}}_{nj}^{\hat{G}}$, we then have

$$\tilde{\mathbf{A}}_{nj}(\mathbf{b}_j, \mathbf{B}_j) = n^{-1} \sum_{i=1}^n \frac{\delta_{ji} \mathbf{Z}_i^{\otimes 2}}{\hat{G}_j(Y_{ji}) \sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \mathbf{b}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \right\},$$

and

$$\tilde{\mathbf{P}}_{nj}^{\hat{G}}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{B}_j) = n^{-1} \sum_{i=1}^n \frac{\mathbf{Z}_i^{\otimes 2} \delta_1 \delta_2 I\{g_{j^*}^{-1}(Y_{j^*i}) \leq \mathbf{Z}_i^T \mathbf{b}_{j^*}\}}{\hat{G}(Y_{1i}, Y_{2i}) \sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \mathbf{b}_j - g_j^{-1}(Y_{ji})}{\sqrt{\mathbf{Z}_i^T \mathbf{B}_j \mathbf{Z}_i}} \right\},$$

for $j^* = 3 - j$.

In the proposed covariance estimation procedure, we estimate $\mathbf{A}_j(\mathbf{b}_j)$ and $\mathbf{P}_j(\mathbf{b}_1, \mathbf{b}_2)$ via smoothing functions $\tilde{\mathbf{A}}_{nj}(\mathbf{b}_j, \mathbf{B}_j)$ and $\tilde{\mathbf{P}}_{nj}^{\hat{G}}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{B}_j)$. Therefore, to show the consistency of $\hat{\psi}_i(\boldsymbol{\tau})$, it is sufficient to show that

$$\tilde{\mathbf{A}}_{nj}(\boldsymbol{\beta}_{j0}, \mathbf{B}_j) \xrightarrow{p} \mathbf{A}_j(\boldsymbol{\beta}_{j0}) \text{ and } \tilde{\mathbf{P}}_{nj}^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_j) \xrightarrow{p} \mathbf{P}_j(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}),$$

for any \mathbf{B}_j satisfying $\mathbf{B}_j = O(n^{-1})$ and $\text{eigmin}(\mathbf{B}_j) > 0$, with the consistency of $\hat{\boldsymbol{\beta}}_j(\tau_j)$, $\hat{G}_j(\cdot)$ and $\hat{G}(\cdot)$. For the consistency of $\tilde{\mathbf{A}}_{nj}(\boldsymbol{\beta}_{j0}, \mathbf{B}_j)$, the proof has been established by [Pang et al. \(2012\)](#) in the proof of Theorem 1(ii). We then only need to show the consistency of $\tilde{\mathbf{P}}_{nj}^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_j)$. Without loss of generality, we treat $j = 1$ and prove $\tilde{\mathbf{P}}_{n1}^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_1) \xrightarrow{p} \mathbf{P}_1(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20})$. Let

$$\tilde{\mathbf{P}}_{n1}^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_1) = n^{-1} \sum_{i=1}^n \frac{\mathbf{Z}_i^{\otimes 2} \delta_1 \delta_2 I\{g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G(Y_{1i}, Y_{2i}) \sqrt{\mathbf{Z}_i^T \mathbf{B}_1 \mathbf{Z}_i}} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sqrt{\mathbf{Z}_i^T \mathbf{B}_1 \mathbf{Z}_i}} \right\}.$$

With condition C2 and the consistency of \hat{G} , we have

$$\sup_{\mathbf{b}_1, \mathbf{b}_2, \boldsymbol{\gamma}; \boldsymbol{\tau} \in \mathbb{D}} \|\tilde{\mathbf{P}}_{n1}^{\hat{G}}(\mathbf{b}_1, \mathbf{b}_2, \mathbf{B}_1) - \tilde{\mathbf{P}}_{n1}^G(\mathbf{b}_1, \mathbf{b}_2, \mathbf{B}_1)\| = o_p(1), \quad (\text{F.1})$$

which implies that $\tilde{\mathbf{P}}_{n1}^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_1) - \tilde{\mathbf{P}}_{n1}^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_1) \xrightarrow{p} o_p(1)$. To complete the proof, we then want to show that

$$\tilde{\mathbf{P}}_{n1}^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_1) \xrightarrow{p} \mathbf{P}_1(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}).$$

Let $\sigma_i = \sqrt{\mathbf{Z}_i^T \mathbf{B}_1 \mathbf{Z}_i}$ and $\tilde{\mathbf{P}}_{n1}^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_1)$ be written as

$$\tilde{\mathbf{P}}_{n1}^G(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_1) = n^{-1} \sum_{i=1}^n \frac{\mathbf{Z}_i^{\otimes 2} \delta_1 \delta_2 I\{g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G(Y_{1i}, Y_{2i}) \sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sigma_i} \right\}.$$

Following the arguments in [Pang et al. \(2012\)](#) and [Li et al. \(2014\)](#), the proof can be accomplished by verifying two conditions,

$$\begin{aligned} E \left[\frac{\delta_1 \delta_2 I\{g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G(Y_{1i}, Y_{2i}) \sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sigma_i} \right\} \middle| \mathbf{Z}_i \right] \\ \rightarrow h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}), \end{aligned} \quad (\text{F.2})$$

and

$$\text{Var} \left[\frac{\delta_1 \delta_2 I\{g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G(Y_{1i}, Y_{2i}) \sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sigma_i} \right\} \middle| \mathbf{Z}_i \right] = O(n^{1/2}). \quad (\text{F.3})$$

We begin with

$$\begin{aligned}
& E \left[\frac{\delta_{1I} \delta_{2I} I\{g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G(Y_{1i}, Y_{2i}) \sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sigma_i} \right\} \middle| \mathbf{Z}_i \right] \\
&= E \left(E \left[\frac{I(T_{1i} \leq C_{1i}) I(T_{2i} \leq C_{2i}) I\{g_2^{-1}(T_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G(T_{1i}, T_{2i}) \sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(T_{1i})}{\sigma_i} \right\} \middle| T_{1i}, T_{2i}, \mathbf{Z}_i \right] \middle| \mathbf{Z}_i \right) \\
&= E \left(E \left[\frac{I\{g_2^{-1}(T_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{\sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(T_{1i})}{\sigma_i} \right\} \middle| T_{1i}, T_{2i}, \mathbf{Z}_i \right] \middle| \mathbf{Z}_i \right) \\
&= E \left[\frac{I\{g_2^{-1}(T_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{\sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(T_{1i})}{\sigma_i} \right\} \middle| \mathbf{Z}_i \right] \\
&= \int_{t_1} \int_{t_2} \frac{I\{t_2 \leq g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20})\}}{\sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(t_1)}{\sigma_i} \right\} h_{12}(t_1, t_2 | \mathbf{Z}_i) dt_2 dt_1 \\
&= \int_{t_1} \frac{h_1(t_1, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i)}{\sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(t_1)}{\sigma_i} \right\} dt_1,
\end{aligned}$$

where $h_{12}(t_1, t_2 | \mathbf{Z}) = \partial^2 H(t_1, t_2 | \mathbf{Z}) / \partial t_1 \partial t_2$ and $h_1(t_1, t_2 | \mathbf{Z}) = \partial H(t_1, t_2 | \mathbf{Z}) / \partial t_1$. Via the variable transformation, we let $x = \{g_1^{-1}(t_1) - \mathbf{Z}_i^T \boldsymbol{\beta}_{10}\} / \sigma_i$ and have that

$$\begin{aligned}
& \int_{t_1} \frac{h_1\{t_1, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\}}{\sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(t_1)}{\sigma_i} \right\} dt_1 \\
&= \int_x \frac{h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\}}{\sigma_i} \phi(-x) g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x) \sigma_i dx \\
&= \int_x h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} \phi(-x) g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x) dx \\
&= \int_x h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}) \phi(-x) dx \\
&\quad + \int_x \phi(-x) \left[h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x) \right. \\
&\quad \left. - h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}) \right] dx \\
&= h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}) + (III).
\end{aligned}$$

By the assumption that $g_1(\cdot)$ is a monotone function, there exists s_1 and s_x such that $s_1 = g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10})$ and $s_1 + \sigma_i s_x = g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x)$. Since $\partial H\{g_1(u), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}\} / \partial u = h_1\{g_1(u), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(u)$, we express that $h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x) = h_1\{s_1 + \sigma_i s_x, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\}$, and $h_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}), g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20}) | \mathbf{Z}_i\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10}) = h_1\{s_1,$

$g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20})|\mathbf{Z}_i\}$. Coupled with Condition C5(ii) that $\partial h_1(t_1, t_2|\mathbf{Z})/\partial t_1$ are continuously differentiable with bounded derivatives, we obtain that

$$\begin{aligned}
|(III)| &= \left| \int_x \phi(-x) [h_1\{s_1 + \sigma_i s_x, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20})|\mathbf{Z}_i\} - h_1\{s_1, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20})|\mathbf{Z}_i\}] dx \right| \\
&= \left| \int_x \phi(-x) \frac{h_1\{s_1 + \sigma_i s_x, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20})|\mathbf{Z}_i\} - h_1\{s_1, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20})|\mathbf{Z}_i\}}{\sigma_i s_x} \sigma_i s_x dx \right| \\
&= \sigma_i \left| \int_x \phi(-x) s_x \partial h_1\{t_1, g_2(\mathbf{Z}_i^T \boldsymbol{\beta}_{20})|\mathbf{Z}_i\} / \partial t_1 \Big|_{t_1=s_1+\sigma_i s_x} dx \right| \\
&\leq M \sigma_i \int_x \phi(-x) |s_x| dx = O(n^{-1/2}),
\end{aligned}$$

where M is the upper bound for $\partial h_1(t_1, t_2)/\partial t_1$. Therefore, the (F.2) condition is satisfied.

For the (F.3) condition, with the assumption on the censoring in Condition C2 and the boundedness of $f_1(t|z)$ in Condition C3(i), it is straightforward to show that

$$\begin{aligned}
&\text{Var} \left[\frac{\delta_{1i} \delta_{2i} I\{g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G(Y_{1i}, Y_{2i}) \sigma_i} \phi \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sigma_i} \right\} \middle| \mathbf{Z}_i \right] \\
&\leq E \left[\frac{\delta_{1i} \delta_{2i} I\{g_2^{-1}(Y_{2i}) \leq \mathbf{Z}_i^T \boldsymbol{\beta}_{20}\}}{G^2(Y_{1i}, Y_{2i}) \sigma_i^2} \phi^2 \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sigma_i} \right\} \middle| \mathbf{Z}_i \right] \\
&< \delta^{-1} \sigma_i^{-2} E \left[\phi^2 \left\{ \frac{\mathbf{Z}_i^T \boldsymbol{\beta}_{10} - g_1^{-1}(Y_{1i})}{\sigma_i} \right\} \middle| \mathbf{Z}_i \right] \\
&= \delta^{-1} \sigma_i^{-2} \int_x \phi^2(-x) f_1\{g_1(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x)|\mathbf{z}\} g_1'(\mathbf{Z}_i^T \boldsymbol{\beta}_{10} + \sigma_i x) \sigma_i dx \\
&\leq \delta^{-1} \sigma_i^{-1} M_{f_1} \int_x \phi^2(-x) dx = O(n^{1/2}),
\end{aligned}$$

where M_{f_1} is the upper bound for $\sup_{t_1, \mathbf{z}} f_1(t_1|\mathbf{z})$. By the results in (F.1) and the conditions (F.2) and (F.3), we complete the proof of the consistency of $\tilde{\mathbf{P}}_{nj}^{\hat{G}}(\boldsymbol{\beta}_{10}, \boldsymbol{\beta}_{20}, \mathbf{B}_j)$, which implies that $\hat{\psi}_i(\boldsymbol{\tau})$ is a consistent estimator for $\psi_i(\boldsymbol{\tau})$.

BIBLIOGRAPHY

- Anderson, J. E., Louis, T. A., Holm, N. V. and Harvald, B. (1992) Time-dependent association measures for bivariate survival distributions. *Journal of the American Statistical Association*, **87**, 641–650.
- ARDES-Group (1999) The age-related eye disease study (areds): design implications areds report no. 1. *Controlled clinical trials*, **20**, 573.
- Bang, H. and Tsiatis, A. (2002) Median regression with censored cost data. *Biometrics*, **58**, 643–649.
- Beyersmann, J., Allignol, A. and Schumacher, M. (2012) *Competing risks and multistate models with R*. New York: Springer.
- Brown, B. and Wang, Y. (2005) Standard errors and covariance matrices for smoothed rank estimators. *Biometrika*, **92**, 149–158.
- (2007) Induced smoothing for rank regression with censored survival times. *Statistics in medicine*, **26**, 828–836.
- Cheng, Y., Fine, J. P. and Kosorok, M. R. (2009) Nonparametric association analysis of exchangeable clustered competing risks data. *Biometrics*, **65**, 385–393.
- Clayton, D. (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–151.
- Cox, D. R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 187–220.
- Fine, J. P. and Gray, R. J. (1999) A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496–509.
- Gerds, T. A., Scheike, T. H. and Andersen, P. K. (2012) Absolute risk regression for competing risks: interpretation, link functions, and prediction. *Statistics in Medicine*, **31**, 3921–3930.
- Gill, R. D. (1980) Censoring and stochastic integrals. *Mathematical Centre Tracts 124*.

- Goldberg, Y. and Kosorok, M. R. (2012) Q-learning with censored data. *Annals of Statistics*, **40**, 529–560.
- Guo, X. and Tsiatis, A. (2005) A weighted risk set estimator for survival distributions in two-stage randomization designs with censored survival data. *The International Journal of Biostatistics*, **1**, 1–17.
- Henderson, R., Ansell, P. and Alshibani, D. (2010) Regret-regression for optimal dynamic treatment regimes. *Biometrics*, **66**, 1192–1201.
- Hernan, M. A. and Robins, J. M. (2010) *Causal inference*. CRC Boca Raton, FL:.
- Hsu, L. and Prentice, L. (1996) On assessing the strength of dependency between failure time variates. *Biometrika*, **83**, 491 – 506.
- Hu, T. and Nan, B. (2011) Time-dependent cross ratio estimation for bivariate failure times. *Biometrika*, **98**, 341–354.
- Huang, Y. (2002) Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association*, **97**, 318–327.
- Ji, S., Peng, L., Li, R. and Lynn, M. J. (2014) Analysis of dependently censored data based on quantile regression. *Statistica Sinica*, **24**, 1411.
- Klein, J. P. and Andersen, P. K. (2005) Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*, **61**, 223–229.
- Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker, R. et al. (2008) Censored quantile regression redux. *Journal of Statistical Software*, **27**, 1–25.
- Lai, T. and Ying, Z. (1988) Stochastic integrals of empirical-type processes with applications to censored regression. *Journal of multivariate analysis*, **27**, 334–358.
- Lakhal, L., Rivest, L.-P. and Beaudoin, D. (2009) IPCW estimator for Kendall’s tau under bivariate censoring. *The International Journal of Biostatistics*, **5**, 10.2202/1557–4679.1121.
- Li, R., Cheng, Y., Chen, Q. and Fine, J. (2016) Quantile association for bivariate survival data. *Biometrics*, 10.1111/biom.12584.
- Li, R., Cheng, Y. and Fine, J. P. (2014) Quantile association regression models. *Journal of the American Statistical Association*, **109**, 230–242.
- Li, R. and Peng, L. (2011) Quantile regression for left-truncated semicompeting risks data. *Biometrics*, **67**, 701–710.
- (2015) Quantile regression adjusting for dependent censoring from semicompeting risks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 107–130.

- Lunceford, J. K., Davidian, M. and Tsiatis, A. A. (2002) Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics*, **58**, 48–57.
- Matthay, K. K., Reynolds, C. P., Seeger, R. C., Shimada, H., Adkins, E. S., Haas-Kogan, D., Gerbing, R. B., London, W. B. and Villablanca, J. G. (2009) Long-term results for children with high-risk neuroblastoma treated on a randomized trial of myeloablative therapy followed by 13-cis-retinoic acid: a children’s oncology group study. *Journal of Clinical Oncology*, **27**, 1007–1013.
- Miyahara, S. and Wahed, A. S. (2010) Weighted kaplan–meier estimators for two-stage treatment regimes. *Statistics in Medicine*, **29**, 2581–2591.
- Murphy, S. A. (2003) Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**, 331–355.
- Nan, B., Lin, X., Lisabeth, L. D. and Harlow, S. D. (2006) Piecewise constant cross-ratio estimation for association of age at a marker event and age at menopause. *Journal of the American Statistical Association*, **101**, 65–77.
- Oakes, D. (1982) A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B: Methodological*, **44**, 414–422.
- (1989) Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 487–493.
- (2008) On consistency of Kendall’s tau under censoring. *Biometrika*, 997–1001.
- Orellana, L., Rotnitzky, A. and Robins, J. M. (2010) Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *The international journal of biostatistics*, **6**.
- Pang, L., Lu, W. and Wang, H. J. (2012) Variance estimation in censored quantile regression via induced smoothing. *Computational statistics & data analysis*, **56**, 785–796.
- Peng, L. and Fine, J. (2009) Competing risks quantile regression. *Journal of the American Statistical Association*, **104**, 1440–1453.
- Peng, L. and Huang, Y. (2008) Survival analysis with quantile regression models. *Journal of the American Statistical Association*, **103**, 637–649.
- Portnoy, S. (2003) Censored regression quantiles. *Journal of the American Statistical Association*, **98**, 1001–1012.
- Prentice, R. and Cai, J. (1992) Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika*, **79**, 495–512.

- Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, **66**, 688.
- Scheike, T. H., Zhang, M.-J. and Gerds, T. A. (2008) Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, **95**, 205–220.
- Shih, J. and Louis, T. (1995) Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 1384–1399.
- Swaroop, A., Chew, E. Y., Bowes Rickman, C. and Abecasis, G. R. (2009) Unraveling a multifactorial late-onset disease: from genetic susceptibility to disease mechanisms for age-related macular degeneration. *Annual review of genomics and human genetics*, **10**, 19–43.
- Tang, X. and Wahed, A. S. (2015a) Cumulative hazard ratio estimation for treatment regimes in sequentially randomized clinical trials. *Statistics in Biosciences*, **7**, 1–18.
- (2015b) Pattern-mixture-type estimation and testing of neuroblastoma treatment regimes. *Journal of Statistical Theory and Practice*, **9**, 266–287.
- Tsiatis, A. (2007) *Semiparametric theory and missing data*. Springer Science & Business Media.
- Wahed, A. S. and Tsiatis, A. A. (2006) Semiparametric efficient estimation of survival distributions in two-stage randomisation designs in clinical trials with censored data. *Biometrika*, **93**, 163–177.
- Wang, W. and Wells, M. T. (2000) Estimation of Kendall’s tau under censoring. *Statistica Sinica*, **10**, 1199–1215.
- Wang, Y., Shao, Q. and Zhu, M. (2009) Quantile regression without the curse of unsmoothness. *Computational Statistics & Data Analysis*, **53**, 3696–3705.
- Wei, L., Ying, Z. and Lin, D. (1990) Linear regression analysis of censored survival data based on rank tests. *Biometrika*, **77**, 845–851.
- Yan, J. and Fine, J. (2005) Functional association models for multivariate survival processes. *Journal of the American Statistical Association*, **100**, 184–196.
- Yavuz, I., Chng, Y. and Wahed, A. S. (2018) Estimating the cumulative incidence function of dynamic treatment regimes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **181**, 85–106.
- Yin, G. and Cai, J. (2005) Quantile regression models with multivariate failure time data. *Biometrics*, **61**, 151–161.
- Zeng, D., Chen, Q. and Ibrahim, J. G. (2009) Gamma frailty transformation models for multivariate survival times. *Biometrika*, **96**, 277–291.

Zhao, Y., Kosorok, M. R. and Zeng, D. (2009) Reinforcement learning design for cancer clinical trials. *Statistics in Medicine*, **28**, 3294–3315.